

Multiple Indicator Growth Mixture Models: eine statistische Simulation zur Performanzevaluation für sozialwissenschaftliche Analysen

Krause, Thomas

Veröffentlichungsversion / Published Version

Dissertation / phd thesis

Empfohlene Zitierung / Suggested Citation:

Krause, T. (2019). *Multiple Indicator Growth Mixture Models: eine statistische Simulation zur Performanzevaluation für sozialwissenschaftliche Analysen*. Stuttgart. <https://doi.org/10.18419/opus-10420>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

**Multiple Indicator Growth Mixture Models: eine
statistische Simulation zur Performanzevaluation für
sozialwissenschaftliche Analysen.**

Von der Fakultät Wirtschafts- und Sozialwissenschaften der
Universität Stuttgart zur Erlangung der Würde eines Doktors
der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.)
genehmigte Abhandlung

Vorgelegt von
Thomas Krause
aus Waiblingen

Hauptberichter: Prof. Dr. Dieter URBAN
Mitberichter: Prof. Dr. Jochen MAYERL

Tag der mündlichen Prüfung: 13.05.2019

Institut für Sozialwissenschaften der Universität Stuttgart

Inhaltsverzeichnis

Tabellenverzeichnis	6
Abbildungsverzeichnis	8
Abstract - deutsch	11
Abstract - english	12
1 Einleitung	13
1.1 Aufbau der Arbeit	13
1.2 Relevanz von Längsschnittstudien und GMMs	14
1.3 Latente Wachstumskurvenmodelle	15
2 Idee von GMMs	19
2.1 Prinzip und Schätzung von GMMs	21
2.1.1 Prinzip des EM-Algorithmus	23
2.1.2 Modellkonstruktion	24
2.1.3 Identifikation der Anzahl latenter Klassen	26
2.2 GMMs als MIGMMs für die Sozialforschung	31
3 Bisherige Ergebnisse	33
3.1 Überblick zu empirischen Anwendungsbereichen von (MI)GMMs	33
3.2 Überblick zu bisherigen Ergebnissen zur Performanz	37
3.2.1 Maße der Klassenidentifikation	37
3.2.2 Klassenseparation und Klassenanteil	38
3.2.3 Konditionale Normalität der Daten	40
3.2.4 Fallzahl	40
3.2.5 Kovariate und GMMs	41
3.2.6 Spezifikationsfehler und Invarianzannahmen	42
4 Offene Fragen zu MIGMMs	44
4.1 Desiderate bzgl. Performanz von MIGMMs	45
4.2 Forschungsfragen und Relevanz der Studie	47

5	Prinzipien für Monte Carlo Studien	49
5.1	Grundidee von Monte Carlo Studien	49
5.2	Durchführung und Darstellung von MC-Studien	51
5.3	Konzept der Umsetzung der MC-Studie	53
6	Thesen/Annahmen bzgl. variierter Modellparameter	54
6.1	Exkurs: Wachstumsmodelle und Messinvarianzen	54
6.2	Thesen/Erwartungen bzgl. Messinvarianzen in MIGMMs	57
6.3	Thesen/Annahmen bzgl. weiterer Modellgrößen	59
7	Konfiguration der Simulation	60
7.1	Zielgrößen/Abhängige Variablen	60
7.2	Konstante Modellparameter	63
7.3	Variierte Modellparameter	66
7.3.1	Manipulierte Messinvarianzen	66
7.3.2	Mixing Proportions	69
7.3.3	Fallzahl	70
7.3.4	Klassenseparation	70
7.3.5	Zusammenfassung bzgl. variierender Größen	71
8	Simulationskonfiguration und Populationsparameter	73
8.1	Modellannahmen von MI/SO-Wachstumsmodellen	74
8.2	Ausgangsmodell	75
8.2.1	Populationswerte Wachstumsverlauf	76
8.2.2	Populationswerte Fehlervarianzen	77
8.2.3	Tabellarische Darstellung der Invarianzen von Intercepts und Faktorladungen	78
8.3	Schätzung der Simulationsdaten	79
8.4	Zufallsziehung	83
8.5	Mixture-Model-Simulationen und Labelswitching	83
8.6	Exkurs: EDV-technische Umsetzung der Simulationsberechnungen	84
9	Ergebnisse zur Klassenidentifikation	86
9.1	Haupteffekte bzgl. AIC, BIC, aBIC	86
9.2	Interaktionseffekte bzgl. AIC, BIC, aBIC	92

9.3	Modellbasierte Analyse bzgl. AIC, BIC, aBIC	99
9.3.1	Modellbasierte Haupteffekte	101
9.3.2	Modellbasierte Interaktionseffekte	104
9.4	Diskussion zur IC-Klassenidentifikation	106
9.5	Zusatzsimulation: Ergebnisse bzgl. BLRT	113
9.5.1	Haupteffekte bzgl. BLRT	114
9.5.2	Modellbasierte Analysen bzgl. BLRT	116
9.5.3	Diskussion zur BLRT-Klassenidentifikation	119
10	Ergebnisse zu Parameterschätzern	122
10.1	Haupteffekte bzgl. des Bias	123
10.2	Modellbasierte Analyse bzgl. des Bias	130
10.2.1	Analyse bzgl. des Bias Ko-/Varianzparameter	131
10.2.2	Analyse bzgl. des Bias der Interceptfaktor-Mittelwerte	135
10.2.3	Analyse bzgl. des Bias der Slopefaktor-Mittelwerte	138
10.3	Diskussion zu den modellbasierten Analysen bzgl. des Bias	144
10.4	Exkurs: Bias der geschätzten Klassenanteile	148
11	Ergebnisse bzgl. der Standardfehler	152
11.1	Haupteffekte bzgl. SE-Verzerrungen	153
11.2	Modellbasiert Analyse bzgl. des SE-Bias	157
11.2.1	Analyse bzgl. des SE-Bias der Ko-/Varianzparameter	157
11.2.2	Analyse bzgl. des SE-Bias der Interceptfaktor-Mittelwerte	160
11.2.3	Analyse bzgl. des SE-Bias der Slopefaktor-Mittelwerte	163
11.3	Diskussion zu den modellbasierten Analysen bzgl. des SE-Bias	166
12	Zusammenfassung der Ergebnisse	170
13	Diskussion der Leistungsfähigkeit von MIGMMs für sozialwissenschaftliche Analysen	174
13.1	Desiderate für nachfolgende Analysen	176
13.2	Resümee	179
14	Literatur	180
A	Anhang	191

A.1	Tabellen und Abbildungen zur Klassenidentifikation	191
A.2	Relativer Bias Referenz	200
A.3	Mediatoranalysen zum Bias	203
A.4	Relativer SE-Bias Referenz	210
A.5	Online-Appendix	213

Tabellenverzeichnis

1	Übersicht konstanter und varianter Simulationsgrößen	72
2	Populationswerte der Wachstumsparameter der Simulation	76
3	Skalierte Populationswerte der Fehlervarianzen der latenten Faktoren (ζ) und manifesten Variablen (ϵ).	78
4	Populationswerte der Faktorladungen für alle untersuchten Invarianzformen . . .	80
5	Populationswerte der Intercepts für alle untersuchten Invarianzformen	81
6	Prozentuale Klassenidentifikation nach Invarianz-Stufen	86
7	Prozentuale Klassenidentifikation nach Klassendistanz-Stufen	89
8	Prozentuale Klassenidentifikation nach Nobs-Stufen	90
9	Prozentuale Klassenidentifikation nach Mixing-Stufen	90
10	Zusammenfassung der IC-Performanz bzgl. der korrekten Klassenidentifikation .	112
11	Proz. BLRT-Klassenidentifikation nach Invarianz-Stufen	114
12	Proz. BLRT-Klassenidentifikation nach Sep-Stufen	115
13	Proz. BLRT-Klassenidentifikation nach Nobs-Stufen	115
14	Proz. BLRT-Klassenidentifikation nach Mixing-Stufen	116
15	Zusammenfassung der BLRT-Performanz bzgl. der korrekten Klassenidentifikation	121
16	Relativer Bias nach Invarianz-Stufen	124
17	Relativer Bias nach Klassendistanz-Stufen	128
18	Relativer Bias nach Fallzahl-Stufen	129
19	Relativer Bias nach Mixing-Stufen	130
20	Regressions-Ergebnisse zu Bias der Ko-/Varianz Parameter (robuste SE; HC3) .	132
21	Delta-R-Quadrat der Praediktoren zum Bias der Ko-/Varianz Parameter	133
22	Regressions-Ergebnisse zum Bias der Interceptfaktoren (robuste SE; HC3) . . .	139
23	Delta-R-Quadrat der Praediktoren zum Bias der Intercept Parameter	140
24	Delta-R-Quadrat der Praediktoren zum Bias der Slope Parameter	142
25	Regressions-Ergebnisse zum Bias der Slopefaktoren (robuste SE; HC3)	143
26	Zusammenfassung der Bias-Performanz bzgl. der Wachstumsfaktoren	147
27	Korrelationsmatrix Bias-Werte	150
28	Reg.-Ergebnisse zu durchschnittlichen Klassenverzerrungen (robuste SE; HC3)	151
29	Delta-R-Quadrat der Praediktoren der geschätzten Klassenteilsverzerrungen . . .	151
30	Relativer SE-Bias nach Invarianz-Stufen	154

31	Relativer SE-Bias nach Klassendistanz-Stufen	155
32	Relativer SE-Bias nach Fallzahl-Stufen	156
33	Relativer SE-Bias nach Mixing-Stufen	156
34	Regressions-Ergebnisse zu SE-Bias der Ko-/Varianz Parameter (robuste SE; HC3)	158
35	Delta-R-Quadrat der Praediktoren zum SE-Bias der Ko-/Varianz Parameter . .	159
36	Delta-R-Quadrat der Praediktoren zum SE-Bias der Intercept Parameter	161
37	Regressions-Ergebnisse zum SE-Bias der Interceptfaktoren (robuste SE; HC3) . .	162
38	Delta-R-Quadrat der Praediktoren zum SE-Bias der Slope Parameter	164
39	Regressions-Ergebnisse zum SE-Bias der Slopefaktoren (robuste SE; HC3)	165
40	Zusammenfassung der SE-Bias-Performanz bzgl. der Wachstumsfaktoren	169
41	Prozentuale Klassenidentifikation nach allen Konfigurationsparametern (Referenz).	192
42	(Fortsetzung) Prozentuale Klassenidentifikation nach allen Konfigurationspara- metern (Referenz).	193
43	Logitkoeffizienten der Modelle zur Klassenidentifikation nach AIC, BIC, aBIC . .	194
44	Odds-Ratios der Modelle zur Klassenidentifikation nach AIC, BIC, aBIC	195
45	Logitkoeffizienten der Modelle zur Klassenidentifikation nach BLRT	196
46	Odds-Ratios der Modelle zur Klassenidentifikation nach BLRT	197
47	Referenztabelle: Relativer Bias für alle Simulationskonfigurationen (Teil 1) . . .	201
48	Referenztabelle: Relativer Bias für alle Simulationskonfigurationen (Teil 2) . . .	202
49	Reg. Erg. zu Bias der Ko-/Varianz Parameter Mediation (robuste SE; HC3) . . .	204
50	Delta-R-Quadrat der Praediktoren zum Bias der Ko-/Varianz Parameter Mediation	205
51	Reg. Erg. zum Bias der Interceptfaktoren Mediation (robuste SE; HC3)	206
52	Delta-R-Quadrat der Praediktoren zum Bias der Intercept Parameter Mediation	207
53	Reg. Erg. zum Bias der Slopefaktoren Mediation (robuste SE; HC3)	208
54	Delta-R-Quadrat der Praediktoren zum Bias der Slope Parameter Mediation . .	209
55	Referenztabelle: Relativer SE-Bias fuer alle Simulationsconfigurationen (Teil 1) .	211
56	Referenztabelle: Relativer SE-Bias fuer alle Simulationsconfigurationen (Teil 2) .	212

Abbildungsverzeichnis

1	Rohdaten aus Suchanfrage im Web of Science	14
2	Heuristische Grundstruktur eines LGC-Modells mit einer linearen Steigung und 4 Erhebungszeitpunkten.	17
3	Grundstruktur eines GMM-Modells mit einer linearen Steigung.	22
4	Grundstruktur eines MIGMM-Modells mit einer linearen Steigung.	34
5	Durchschnittliches Verlaufsmuster der latenten Wachstumsklassen nach den Ab- stufungen der Klassendistanz	77
6	Populationswerte des (messtechnisch invarianten) MIGMM-Modells.	82
7	Interaktionseffekt zwischen Fallzahl und Klassenanteilen (Nobs*Mixing) bzgl. Klassenidentifikation nach AIC, BIC und aBIC.	93
8	Interaktionseffekt zwischen Invarianzabstufungen und Klassendistanz (Invar*Sep) bzgl. Klassenidentifikation nach AIC, BIC und aBIC.	95
9	Interaktionseffekt zwischen Invarianzabstufungen und Fallzahl (Invar*Nobs) bzgl. Klassenidentifikation nach AIC, BIC und aBIC.	97
10	Interaktionseffekt zwischen Invarianzabstufungen und Klassenanteilen (Invar*Mixing) bzgl. Klassenidentifikation nach AIC, BIC und aBIC.	100
11	AMEs bzgl. der abhängigen Variablen der korrekten Klassenidentifikation nach AIC, BIC und aBIC	102
12	AMEs bzgl. der abhängigen Variablen der korrekten Klassenidentifikation nach BLRT.	118
13	Mosaik-Plots zur Klassenidentifikation (aBIC, BIC)	198
14	Mosaik-Plots zur Klassenidentifikation (AIC, BLRT)	199

Abkürzungsverzeichnis

aBIC / SSABIC Sample-Size Adjusted BIC

AIC Akaikes Information Criterion

AMLR (Adjusted) Lo-Mendell-Rubin Test

AME Average Marginal Effects

BIC Bayes Information Criterion

BLRT Bootstrapped Parametric Likelihood Ratio Test

CAIC Consistent Akaikes Information Criterion

CLC Classification Likelihood Criterion

CP Coverage Probabilities

DGP Data Generation Process

EM Expectation-Maximization

GMM Growth Mixture Model

IC Information Criteria

ICL-BIC Integrated-Completed Likelihood Criterion

INV Invarianzstufe (MC-Simulationsvariable)

LCGA Latent Class Growth Analysis

LGC Latent Growth Curve

LL Log-Likelihood

LRT Likelihood-Ratio-Test

MAR Missing at Random

MD Mahalanobis Distance

MC Monte Carlo

MIGMM	Multiple Indicator Growth Mixture Model
MIXING	Klassenanteilsverhältnis (MC-Simulationsvariable)
MSE	Mean Squared Error
NEC	Normalized Entropy Criterion
NOBS	Fallzahl / Number of Observations (MC-Simulationsvariable)
SE	Standard Error
SEP	Klassendistanz (MC-Simulationsvariable)
SCAIC	SCAIC
SO	Second Order
SOGMM	Second Order Growth Mixture Model

Abstract - deutsch

Multiple Indicator Growth Mixture Models: eine statistische Simulation zur Performanzevaluation für sozialwissenschaftliche Analysen.

Multiple Indicator Growth Mixture Modelle (MIGMM) kombinieren die Konstruktionsprinzipien von latenten Messmodellen, Wachstumskurvenmodellen und der latenten Klassenanalyse. MIGMMs sind somit analytische Werkzeuge für die empirische Sozialforschung, welche es ermöglichen eine Post-hoc-Identifikation und Beschreibung von Gruppenunterschieden hinsichtlich des zeitlichen Wandels durchzuführen und gleichzeitig die Messungen als latente Konstrukte zu berücksichtigen. Durch die Identifikation von unbeobachteten Subpopulationen lassen sich soziale Veränderungsprozesse und deren Unterschiede zwischen und innerhalb der unbeobachteten Subpopulationen untersuchen. Während einfache Growth Mixture Modelle schon anhand von zahlreichen Monte Carlo Studien evaluiert wurden, steht eine systematische Analyse der Leistungsfähigkeit von Multiple-Indicator-GMMs noch weitestgehend aus. Diese Simulationsstudie zielt darauf ab, die Performanz von MIGMMs unter verschiedenen Datensituationen systematisch zu evaluieren, wobei insbesondere zeitliche, gruppenspezifische und kombinierte Invarianzverletzungen der latenten Messmodelle im Fokus sind.

Dazu wurden folgende Bedingungen manipuliert: Messinvarianz der latenten Konstrukte, Klassendistanz, Klassenanteilsverhältnisse und die Fallzahl. Es zeigt sich, dass insbesondere die Klassendistanz zwischen den latenten Verlaufsgruppen sowie die Verletzungen der Messinvarianz zwischen den latenten Gruppen einen Einfluss auf die Performanz von MIGMMs hinsichtlich Klassenidentifikation, Koeffizienten-Schätzungen und der Standardfehlerschätzungen haben. Auf Grundlage dieser Ergebnisse wird empfohlen, MIGMMs ausschließlich bei großen und klaren Unterschieden der Verlaufsmuster und nur für latente Konstrukte mit bereits etablierter Zuverlässigkeit hinsichtlich Gruppendifferenzen einzusetzen. Die Ergebnisse dieser Untersuchung sollen dazu dienen, die Forschungspraxis der empirischen Sozialforschung anzuleiten und Empfehlungen zum Einsatz dieser neuen Analysewerkzeuge für längsschnittliche Untersuchungen zu geben.

Abstract - english

Multiple Indicator Growth Mixture Models: a statistical simulation to evaluate performance for social science analysis.

Multiple Indicator Growth Mixture Models (MIGMM) combine the design principles of latent measurement models, growth curve models, and latent class analysis. MIGMMs are thus analytical tools for empirical social research, which consider the measurements as latent constructs and simultaneously allow post-hoc identification and description of group differences with respect to temporal change. By identifying unobserved subpopulations, social change processes and their differences between and within the unobserved subpopulations can be investigated. While simple Growth Mixture Models, based on manifest variables, have already been evaluated in numerous Monte Carlo studies, a systematic analysis of the performance of multiple indicator GMMs is still lacking. This simulation study aims to systematically evaluate the performance of MIGMMs under different data situations, focusing in particular on temporal, group-specific and combined invariance violations of the latent measurement models.

The following conditions were manipulated: measurement invariance of the latent constructs, class distance, mixing proportion and the number of observations. It is shown that especially the class distance between the latent growth groups and the violations of the measurement invariance between the latent groups have an influence on the performance of MIGMMs regarding class identification, coefficient estimates and standard error calculation. Based on these results, it is recommended to use MIGMMs only for large and distinct differences in the growth pattern and only for latent constructs with established reliability for group differences. The results of this study are intended to guide the research practice of empirical social research and provide recommendations for the use of this new analytic tool for longitudinal studies.

1 Einleitung

1.1 Aufbau der Arbeit

Nach einer einleitenden Hinführung zum Thema dieser Arbeit in Kapitel 1 sollen die Kapitel 2 bis 3 dazu dienen, die Idee und Berechnungsprinzipien von sogenannten Growth Mixture Modellen (GMMs) (vgl. Muthen, 2001) sowie deren Erweiterung als Multiple Indicator Growth Mixture Modellen (MIGMMs) zu verdeutlichen. Dabei soll im Weiteren auf die bisherigen empirischen Anwendungsfelder sowie auf bisherige Ergebnisse zur Performanz von GMMs im Rahmen von Simulationsstudien eingegangen werden. Auf Basis dieses Vorgehens soll verdeutlicht werden, inwiefern der Einsatz von GMMs nützlich sein kann, wo sie bereits eingesetzt werden und welche Limitationen und Probleme bei der Schätzung von GMMs bereits bekannt sind.

Aufbauend auf diesen Kapiteln und der darin enthaltenen bisherigen Forschungsergebnissen soll in den darauf folgenden Kapiteln 4 bis 6 auf Forschungsdesiderate bezüglich der Performanz von Multiple Indicator Growth Mixture Modells (MIGMMs) eingegangen werden. Dazu werden Arbeitshypothesen entwickelt, welche auf den bisherigen Analysen zur Leistungsfähigkeit von GMMs sowie auf Ergebnissen hinsichtlich von LGC-Modellen zweiter Ordnung basieren. Des Weiteren soll die Strategie zur Umsetzung der Tests der Arbeitshypothesen im Rahmen einer Monte-Carlo-Simulation expliziert werden, was auch die Explikation und Rechtfertigung der zentralen Simulationsparameter und deren Ausprägungen mit enthalten soll.

In den Kapiteln 7 bis 12 soll schließlich, neben einer detaillierten Darstellung der Simulationskonfiguration und EDV-technischen Umsetzung, der Test der Hypothesen erfolgen und entsprechend dokumentiert werden. Auf Grundlage der Performanz hinsichtlich der Zielgrößen der Monte-Carlo-Simulation können hierbei Schlüsse und Empfehlungen bezüglich der empirischen Forschungspraxis abgeleitet werden. Besonders die korrekte Identifikation von latenten Klassen im Rahmen von MIGMMs sowie deren Konditionalität bzgl. der variierten Simulationsparameter steht bei dieser Auswertung im Fokus. Kapitel 13 schließt mit einer zusammenfassenden Betrachtung und Bewertung der Simulationsergebnisse ab, deren Auswirkungen und Empfehlung für die empirische Forschungspraxis in Abschnitt 13.1 diskutiert werden. Im letzten Abschnitt dieser Arbeit werden schließlich noch neue Forschungsdesiderate aufgezeigt, welche sich aus den Ergebnissen der Analyse ergeben haben.

1.2 Relevanz von Längsschnittstudien und GMMs

Die zunehmende Verfügbarkeit von umfangreichen und verlässlichen Datenbeständen mit Längsschnittdaten in den Sozialwissenschaften sowie der sich immer weiter verbreitende Grundkonsens über die Wichtigkeit von Längsschnittdesigns bilden die empirische und analytische Grundlage für die Analyse von Entwicklungs- und Wachstumsverläufen. Diese Analysen ermöglichen nicht nur erweiterte deskriptive Befunde, sondern auch die explizite Modellierung von theoretisch analytischen Überlegungen und Thesen. Diese Erweiterung in der Forschungspraxis und der empirischen Datengrundlage verlangt auch nach neuen Analysemethoden, welche der Verarbeitung der Informationsbasis gerecht werden, indem die zusätzlich anfallenden Datenmuster adäquat in der Analyse berücksichtigt werden können. Die Modellklasse der sogenannten Growth Mixture Modellen gewinnt in den letzten Jahren für sozialwissenschaftliche Analyse zunehmend an Popularität. Eine einfache Suche im Social Science Citation Index mit dem Begriff «Growth Mixture Model» offenbart, dass sowohl die Anzahl der Publikationen wie auch die Zitation der entsprechenden Literatur seit 2000 stetig zunimmt (vgl. Abbildung 1).

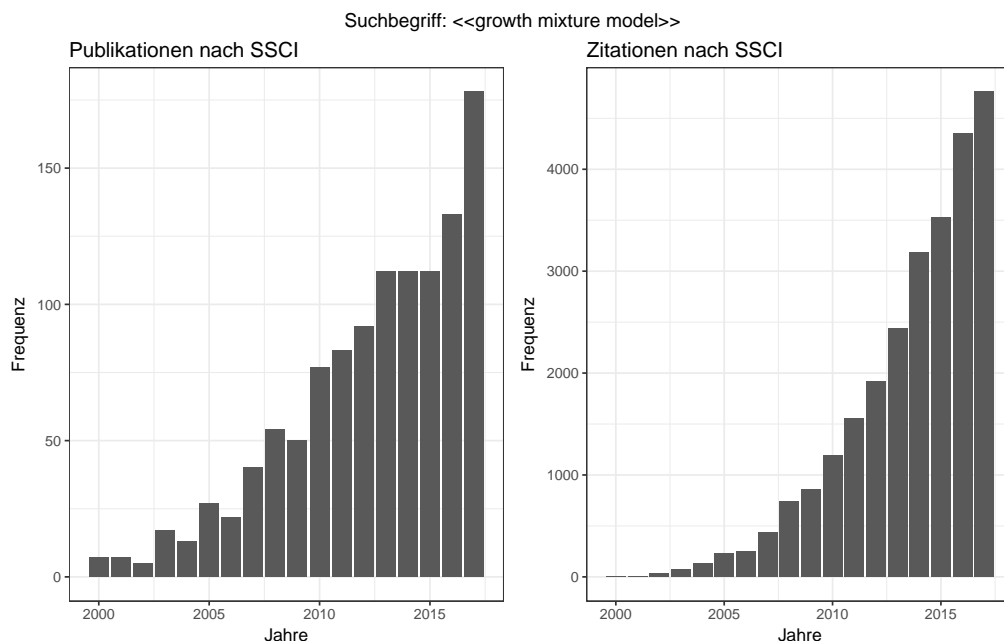


Abbildung 1: Rohdaten aus Suchanfrage im Web of Science

Eigene Darstellung.

Dies liegt wohl nicht zuletzt daran, dass sie neben der dynamischen Entwicklungsperspektive der Längsschnittdaten zudem Gruppendifferenzen in den Blick nehmen. Dies ermöglicht nicht nur die Aufgabe der strikten Annahme von homogenen Entwicklungs-

verlaufen, sondern ermöglicht durch die explorative Komponente dieser Modellklasse zugleich die Entdeckung von nicht a priori festgelegten Entwicklungs- und Verlaufsstrukturen. Ob die Modellklasse der GMMs diesem Anspruch auch gerecht werden kann und welche Verzerrungen und Fehlschlüsse hierbei auftreten können, ist das Thema von einem zunehmenden Korpus der Methodenliteratur zu Strukturgleichungsmodellen. Leider blieb bisher die Kombination von GMMs und latenten Messmodellen, welchen für die Messung im sozialwissenschaftlichen Kontext eine wichtige Rolle zukommt, in Form von Multiple Indicator Growth Mixture Modellen deutlich unterbelichtet. Die vorliegende Arbeit trägt daher dazu bei, dass auch diese Erweiterung für sozialwissenschaftliche Analysen systematisch untersucht wird. Da GMMs als Erweiterung von latenten Wachstumskurvenmodellen verstanden werden können, wird zunächst auf diese Modellklasse eingegangen, bevor die Weiterentwicklungen in Form von GMMs und MIGMMs diskutiert werden.

1.3 Latente Wachstumskurvenmodelle

Latente Wachstumskurvenmodelle (Latent Growth Curve Model; LGCM) sind inzwischen zu einer der dominantesten analytischen Techniken für die Untersuchung von intraindividuellen Veränderungen von Personen, den interindividuellen Unterschieden dieser Veränderungen und den Determinanten dieser Prozesse geworden (vgl. Grimm & Ram, 2009: 124) (siehe auch: McArdle & Epstein, 1987; Meredith & Tisak, 1990; Preacher et al., 2008; Rogosa & Willett, 1985; Urban, 2004).¹

LGC-Modelle ermöglichen es, die Analyse des Wandels einer Variablen über die Zeit zu betrachten. Sie ermöglichen dadurch dynamische Entwicklungsprozesse im Rahmen der Strukturgleichungsmodellierung durchzuführen und modelltechnisch adäquat abzubilden (vgl. Urban, 2004: 1). Dazu werden zeitversetzte wiederholte Messungen einer Variablen als manifeste Indikatoren von latenten Konstrukten (Intercept- und Slope-Konstrukte) verstanden, welche wiederum das Startniveau zu Beginn der Erhebung und den Wandel über die Zeit repräsentieren.

Konzeptuell können LGC-Modelle als ein zweistufiger Prozess verstanden werden. Zunächst wird dabei (sozusagen) der individuelle Verlauf für jede einzelne Analyseeinheit bestimmt (das individuelle Wachstumsverlaufsmuster in Form einer Regressionsgleichung), sodass für jede Einheit eine eigene Verlaufsgleichung, welche durch individuelle

¹Das Wort «Wachstum» darf hierbei nicht wörtlich verstanden werden. LGCM eignen sich zur Analyse von jeder Art von Veränderung, egal ob Wachstum, Verringerung, Stabilität oder jeden Mix aus diesen Verlaufsmustern (vgl. Urban, 2004: 1 f.).

Intercept- und Steigungs-Faktoren charakterisiert ist, aufgestellt werden kann (vgl. Wickrama, 2016: 25). Diese Regressionsgleichung muss nicht notwendiger Weise einem linearen Verlauf folgen. In einem zweiten Schritt werden diese individuelle Verlaufsgleichungen dann über alle Analyseeinheiten zusammengefasst, um ein durchschnittliches Startniveau (oder Sockelbetrag) und eine durchschnittliche Wandlungsrate zu berechnen, welche wiederum eine jeweils eigene Varianz um diese Durchschnittswerte aufweisen (vgl. ebenda). Auch wenn dieser zweistufige Ablauf nicht dem tatsächlichen rechnerischen Vorgehen entspricht, wird dadurch die Idee hinter der Modellierung von latenten Konstrukten in Form von Intercepts und Slopes im Rahmen von LGC-Modellen verdeutlicht. Es wird ein durchschnittliches Verlaufsmuster bestimmt, von welchem die Analyseeinheiten nur insofern abweichen, als dass Varianz bzgl. der Ausgangsniveaus und der Veränderungsraten besteht.

Um dieses Analyseschema im Rahmen einer Strukturgleichungsspezifikation umzusetzen, werden die Intercept- und Slope-Faktoren mit Hilfe von fixierten Faktorladungen mit den zeitversetzten wiederholten Messungen (egal ob diese als manifeste oder latente Konstrukte umgesetzt sind) verbunden. Diese fixierten Faktorladungen können als Zeitkoeffizienten verstanden werden, da sie durch ihre Werte die Zeitachse des Wachstumsprozesses definieren (vgl. Urban, 2004: 77 f.). Abbildung 2 zeigt ein einfaches lineares Wachstumsmodell mit einem Level-/Sockel-/Intercept-Faktor, welches mit jedem Messzeitpunkt über die auf «1» fixierte Faktorladung verbunden ist. In Abbildung 2 ist ebenfalls ein einzelner linearer Steigungsfaktor enthalten, welcher mit den fixierten Faktorladungen 0, 1, 2, 3 mit den (im hier abgebildeten Fall) manifesten Messzeitpunkten verknüpft ist. Durch diese Spezifikation repräsentiert der Intercept-Faktor den Startwert zum ersten Erhebungszeitpunkt, welcher zu jedem darauf folgenden Messzeitpunkt um den Betrag des Slope-Faktors (gewichtet durch die jeweilige Faktorladung) ergänzt wird.² Durch alternative Spezifikationen der Zeitkoeffizienten (bspw. Zeitkoeffizienten des Slope: -2, -1, 0, 1) kann folglich auch die Bedeutung der latenten Wachstumsparameter verschoben werden, was auch zu einer alternativen Bedeutung einer potenziell vorhandenen Kovarianz zwischen Intercept- und Slope-Faktoren führt (vgl. Duncan & Duncan, 2009: 983 ff.). Ebenso kann die Wachstumsverlaufform eines LGC-Modelles durch freie Schätzung einzelner Zeitkoeffizienten oder das Hinzufügen von Polynomen in Form von weiteren Slope-Faktoren ein komplexeres Muster als den linearen Verlauf annehmen (siehe z. B.: Duncan & Dun-

²Diese Form der Spezifikation eines Wachstumsmodells in der SEM-Logik ist konzeptionell eng mit der Logik von Wachstumsmodellen im Rahmen von Mehrebenenmodellen verbunden, welche auf wiederholte Messungen angewendet werden (vgl. hierzu im Detail: Singer & Willett, 2003: 266 ff.; Hox, 2010: 335 f.).

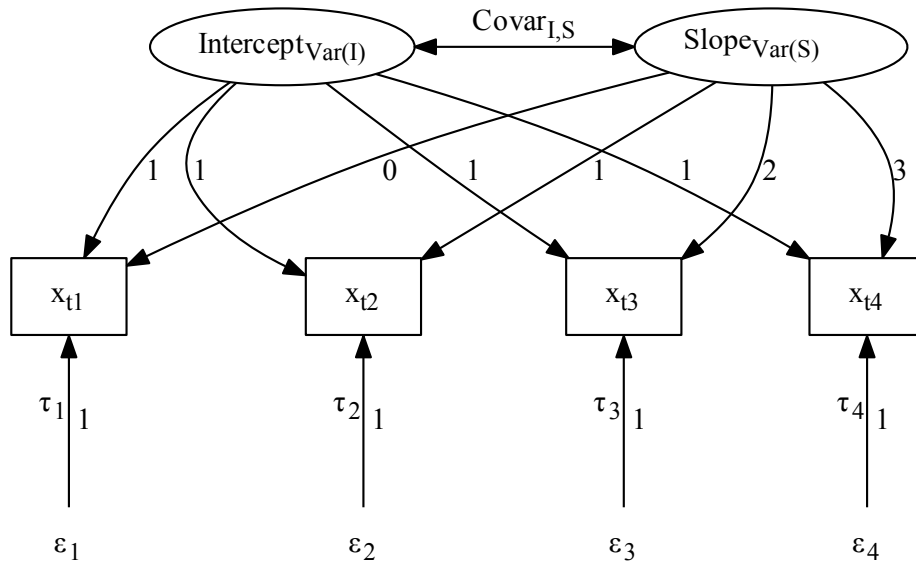


Abbildung 2: Heuristische Grundstruktur eines LGC-Modells mit einer linearen Steigung und 4 Erhebungszeitpunkten.

Die vollständige Mittelwertstruktur ist aus Übersichtlichkeitsgründen nicht abgebildet. Eigene Darstellung.

can, 2009: 985 f.; Urban, 2004: 78 ff. oder 96 ff.).

Die latenten Intercept- und Slope-Faktoren selbst sind über ihre jeweiligen Mittelwerte und ihre Varianzen vollständig charakterisiert. Die Mittelwerte (auch: Mean-Struktur) von Slope und Intercept stehen in der Logik von LGC-Modellierungen für die typischen bzw. durchschnittlichen Werte des Verlaufsmusters der Population. Die Unterschiedlichkeit (oder der Grad der Unterschiedlichkeit) der einzelnen Analyseeinheiten hinsichtlich dieser typischen Verlaufsform drückt sich über die Varianz der Intercept- und Slope-Konstrukte aus. Die so im statistischen Modell abgebildete Unterschiedlichkeit hinsichtlich des Entwicklungsmusters der untersuchten Variable können in einem weiteren Modellierungsschritt über inhaltliche Größen (in Form von Kovariaten) statistisch erklärt werden, wenn Anteile der Varianz der Intercept- und Slope-Konstrukte statistisch gebunden werden können (Conditional LGC-Models).

Dabei unterstellen LGC-Modelle allerdings immer, dass der jeweils modellierte Entwicklungsverlauf (wie auch mögliche Determinanten dieses Verlaufs) für die gesamte untersuchte Gruppe ein gleiches Verlaufsmuster aufweist, von welchem lediglich in Form von zufälligen Fehler abgewichen wird. LGC-Modelle gehen damit davon aus, dass die jeweils

untersuchten Daten aus einer einzigen Population stammen.³ Das jeweilige LGC-Modell ist somit durch ein durchschnittliches Wachstumsverlaufsmuster bestimmt, welches sich wiederum in einer latenten Intercept-Komponente und einer durchschnittlichen latenten Slope-Komponente sowie der individuellen Variation um diese durchschnittlichen Komponenten definiert (vgl. Tolvanen, 2007: 2). Die individuelle Variation um die durchschnittlichen Werte von Intercept und Slope werden hierbei in Form von Varianzen-Parameter berücksichtigt. Formal lässt sich ein einfaches LGC-Modell erster Ordnung über folgende Gleichungen darstellen:

$$\begin{aligned}
y_{ti} &= \lambda_{t1}\eta_{1i} + \dots + \lambda_{tk}\eta_{ki} + \epsilon_{ti} \\
\eta_{1i} &= \alpha_1 + \zeta_{1i} \\
&\dots \\
\eta_{ki} &= \alpha_k + \zeta_{ki}
\end{aligned} \tag{1}$$

Wobei zusätzlich gilt:

$$E[\epsilon] = 0 \ \& \ E[\zeta_k] = 0 \ \& \ COV[\epsilon, \zeta_k] = 0$$

(vgl.: Reinecke, 2012: 6 ff.; Urban, 2004: 78 ff. ; McArdle & Nesselroade, 2003: 455 f.).

Dabei repräsentiert y_{ti} den beobachteten Wert jeder Analyseeinheit i zum Erhebungszeitpunkt t mit einem idiosynkratischen Fehler ϵ_{ti} . λ_{tk} steht für den Zeitskalierungskoeffizienten jedes latenten Wachstumsfaktors k zu jedem Erhebungszeitpunkt t . η_{ki} repräsentiert den Wert jeder Person (i) hinsichtlich eines Wachstumsfaktors k . Dieser Wert auf dem Wachstumsfaktor ist wiederum bestimmt über einen festen Anteil (α_k), welcher den Durchschnitt über alle Analyseeinheiten darstellt und einen variablen Fehleranteil ζ_{ki} , welcher sich über die Analyseeinheiten hinweg (in aller Regel) unterscheidet.

Wird ein LGC-Modell als ein Modell zweiter Ordnung spezifiziert, wobei die Messzeitpunkte nicht durch manifeste Indikatoren, sondern durch latente Konstrukte (welche über mehrere manifesten Indikatoren repräsentiert sind) abgebildet werden (Curve-of-Factors-Form), verschiebt sich die Gleichungsdarstellung um eine zusätzliche Ebene.

³Die Annahme einer singulären Population gilt natürlich nur dann, wenn ein LGC-Modell nicht als Multigruppenmodell spezifiziert wird. In diesem besonderen Fall können auch LGC-Modelle von mehreren zugrundeliegenden Populationen mit einem jeweils eigenen Wachstumsverlauf ausgehen (vgl. Duncan & Duncan, 2009: 987). Allerdings muss bei Multigruppen-LGC-Modellen – im Gegensatz zu GM-Modellen – die Populationszugehörigkeit a priori feststehen.

$$\begin{aligned}
y_{tij} &= \rho_{jt} + \gamma_{jt}\pi_{it} + \epsilon_{tij} \\
\pi_{it} &= \lambda_{t1}\eta_{1i} + \dots + \lambda_{tk}\eta_{ki} + \mu_{it} \\
\eta_{1i} &= \alpha_1 + \zeta_{1i} \\
&\dots \\
\eta_{ki} &= \alpha_k + \zeta_{ki}
\end{aligned} \tag{2}$$

Wobei zusätzlich gilt:

$$\begin{aligned}
E[\epsilon_{tj}] &= 0 \ \& \ E[\mu_t] = 0 \ \& \ E[\zeta_k] = 0 \\
COV[\epsilon_{tj}, \mu_t] &= 0 \ \& \ COV[\zeta_k, \mu_t] = 0 \ \& \ COV[\mu_t, \mu_{t\pm l}] = 0 \ \& \ COV[\epsilon_j, \zeta_k] = 0
\end{aligned}$$

(vgl. auch Urban, 2004: 80 f.). y_{tij} und ϵ_{tij} werden dabei durch den Index j ergänzt, um unterschiedliche Indikatoren eines Messkonstruktes abbilden zu können. Der Term ρ_{jt} steht nun für indikatorspezifische (j) Konstanten über alle Analyseeinheiten zu einem entsprechenden Messzeitpunkt (t). γ_{jt} gibt die Faktorladung zum latenten Messkonstrukt wieder, welche sich bzgl. des manifesten Indikators (j) und hinsichtlich des Messzeitpunktes (t) unterscheiden kann. π_{it} steht schließlich für den Wert auf dem Messfaktor für eine Analyseeinheit (i) zu einem spezifischen Erhebungszeitpunkt (t). Auf der Ebene des latenten Messkonstrukts kommt ein weiterer Fehlerterm μ_{it} hinzu, welcher für Analyseeinheit (i) zu den Erhebungszeitpunkten (t) einen individuellen Wert annehmen kann. Alle weiteren Gleichungsterme sind analog zum Modell erster Ordnung zu verstehen.

2 Idee von GMMs

Growth Mixture Modelle stellen eine Möglichkeit dar, Heterogenität von Populationen explizit in der Analyse zu berücksichtigen statt im vornherein davon auszugehen, dass Entwicklungsverläufe für die gesamte untersuchte Population einem gleichen Verlauf folgen. Die Annahme eines für alle zutreffenden Modells des Wachstumsverlaufs wird dazu fallen gelassen. GMMs erlauben hierfür Parameterunterschiede zwischen unbeobachteten Subpopulationen über latente Wachstumskurvenklassen (vgl. Muthén, 2004: 348). Die dahinterstehende Idee des sogenannten Mixture Modeling ⁴ ist, dass beobachtete Daten

⁴Der Begriff Mixture bezieht sich darauf, dass es sich um eine «Mischung» von unbeobachteten Subpopulationen bzw. um eine Zusammensetzung aus unbeobachteten Gruppen handelt (vgl. Nagin, 2010: 57). Noch genauer ausgedrückt, handelt sich bei der Growth-Mixture-Verteilung um eine (i. d. R.) Mischung

mehrere Subpopulationen bzw. latente Klassen repräsentieren können, welche identifizierbar sind und deren individuelle Parameter anhand der Daten für jede Klasse individuell geschätzt werden können (vgl. Muthen (2001); Tolvanen (2007)). So kann sich, beispielsweise, im Rahmen einer GM-Modellierung zeigen, dass der Wandel einer Einstellungsgröße, welche zuvor anhand eines herkömmlichen LGC-Modells als ein konstanter Verlauf ohne eine durchschnittliche Steigung (aber mit großer Variabilität) zu deuten war, eigentlich grundsätzlich anders ausfallen müsste. So könnte eine GMM zeigen, dass es sich tatsächlich um zwei gänzlich unterschiedlichen Gruppen handelt, welche für sich ein relativ einheitliches Wachstumsverlaufsmuster (z. B. eine Gruppe positiv und die Andere negativ) aufweisen, welches sich allerdings im Durchschnitt der Schätzung des LGC-Modells nur ausgemittelt hat und eine starke Varianz der Wachstumsparameter verursachte. GMMs können somit auch als eine Form von Multigruppenanalyse verstanden werden, in welcher die Klassen allerdings durch die Daten und das jeweilige Analysemodell bestimmt werden und nicht (a priori) explizit durch den/die Forscher/Forscherin selbst (vgl. Hancock & Samuelsen, 2008: 3). Damit stellen GMMs aber ein exploratives und stark an der empirischen Datengrundlage orientiertes Verfahren dar, welches durch die jeweilige Datensituation beeinflusst wird (vgl. Grimm & Ram, 2009: 124). Anders formuliert, stellt die Methodik von GMMs ein analytisches Werkzeug dar, welches es ermöglicht, eine post-hoc Identifikation und Beschreibung von Gruppenunterschieden hinsichtlich des Wandels über die Zeit durchzuführen (vgl. Ram & Grimm, 2009: 565). Damit kann das Problem von unbeobachteter Heterogenität in Wachstumsverläufen zum einen bearbeitet und behoben werden (vgl. Reinecke, 2014: 440) sowie zum anderen auch als explizites analytisches Untersuchungsziel herausgestellt werden. Dieser Unterschied in der methodologisch statistischen Perspektive wird inzwischen auch unter den Stichworten «variable-centered approaches» und «person-centered approaches» diskutiert. Nach dieser Unterscheidung nach Morin et al. (2018) fallen klassische LGC-Modelle unter die «variable-centered» und GM-Modelle unter die «person-centered» Kategorie.

Growth Mixture Modelle bieten sich sowohl zur Analyse von randomisierten Experimentaltstudien wie auch im Kontext von nicht-experimentellen Untersuchungssituationen an. So lassen sich in experimentellen Interventionsstudien nicht nur unterschiedliche Entwick-

von Normalverteilungen von Populationsverlaufsmustern (vgl. Wickrama, 2016: 191). Dabei kann man sich die latenten Klassen als eine Form von Sortierung der Analyseeinheiten aus heterogenen Populationen vorstellen, was wiederum den Oberbegriff der «finite mixture models» plausibilisiert (vgl. Rupp, 2013: 531).

lungsverlaufgruppen basiert auf dem Treatment identifizieren, sondern auch unterschiedliche Verläufe auf Grundlage von unterschiedlichen Wirkungen des Treatments selbst ausmachen (vgl. Muthén & Asparouhov, 2008: 161). Im Ergebnis kann dies zu noch zielgerichteteren Empfehlungen für die empirische Praxis von Interventionen führen, basiert auf jeweilig identifizierten Gruppenzugehörigkeitsmerkmalen. Auch für nicht-experimentelle Untersuchungen kann die Exploration von unterschiedlichen Entwicklungsverläufen ein detaillierteres Bild der Empirie liefern, sowie unterschiedliche Effekte hinsichtlich vorgeordneter Prädiktorgrößen wie auch nachgeordneter Konsequenzen für die einzelnen latenten Entwicklungsgruppen zulassen (vgl. Muthén & Asparouhov, 2008: 161). Zusammenfassend lässt sich also sagen, dass GMMs ein Modellierungsverfahren anbieten, welches es ermöglicht, mehrere unbeobachtete Subpopulationen zu identifizieren, deren Wandel im Längsschnitt zu beobachten und Unterschiede hinsichtlich dieses Wandels zwischen und innerhalb dieser unbeobachteten Subpopulationen zu untersuchen (vgl. Grimm & Ram, 2009: 123). Der Analysefokus richtet sich somit nicht nur auf intraindividuelle Veränderungen oder nur auf interindividuelle Vergleiche, sondern kombiniert beide Perspektiven miteinander. Darüber hinaus ermöglicht das GMM Analyseframework auch die Analyse von Prädiktoren der Zugehörigkeit zu latenten Klassen (und Prädiktoren der Wachstumsparameter) sowie die Untersuchung möglicher Folgen der Zugehörigkeit zu einer latenten Wachstumskurvengruppe. Damit verbinden GMMs die Modellierungstechnik von LGC-Modellen mit der Idee von Mixture Modeling und können als eine Kombination aus beiden Datenanalyseverfahren verstanden werden (vgl. Tolvanen, 2007: 2).

2.1 Prinzip und Schätzung von GMMs

GMMs können als eine Erweiterung von latenten Wachstumskurvenmodellen (LGC-Modellen) verstanden werden. Die Erweiterung besteht hierbei darin, dass nicht nur jeweils ein Vektor an Mittelwerten und eine Kovarianzenmatrix geschätzt wird, sondern dass sich die unterstellten Gruppen des GMMs bzgl. der Mittelwerte und Kovarianzmatrizen jeweils unterscheiden können.

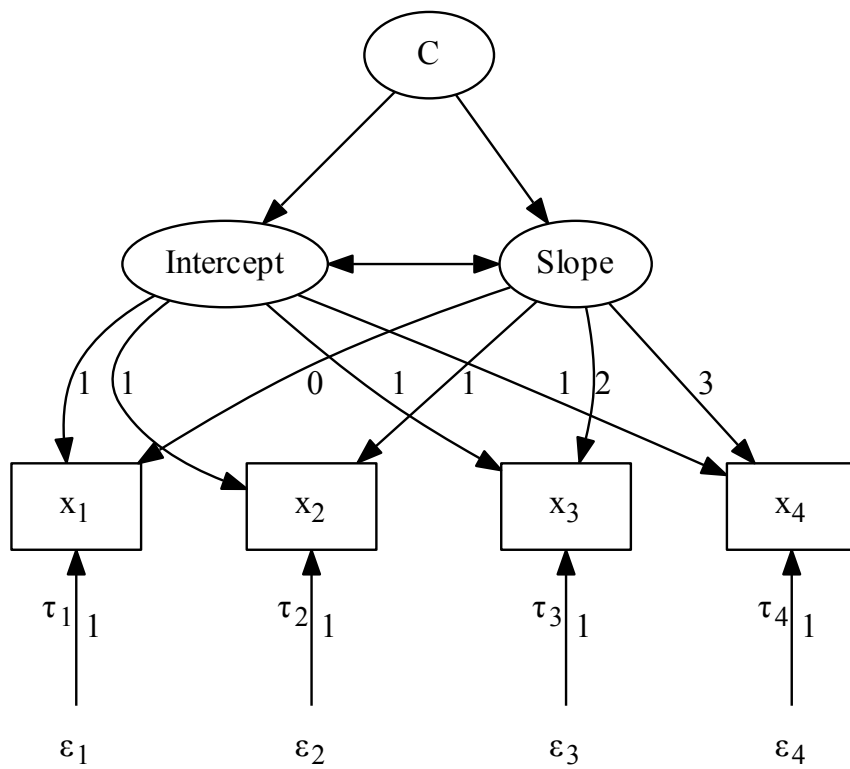


Abbildung 3: Grundstruktur eines GMM-Modells mit einer linearen Steigung.

Diachrone Fehlerkorrelationen und die vollständige Mittelwertstruktur sind aus Übersichtlichkeitsgründen nicht abgebildet. Eigene Darstellung.

Die Wahrscheinlichkeitsdichtefunktion eines GMMs lässt sich durch

$$f(y_i) = \sum_{k=1}^K \pi_k \phi_k[y_i; \mu_k(\theta_k) \Sigma(\theta_k)]$$

Wobei gilt: (3)

$$0 \leq \pi_k \leq 1 \text{ und } \sum_{k=1}^K \pi_k = 1$$

wiedergeben. Dabei stellt $\mu_k(\theta_k)$ die modellimplizierte Mittelwertstruktur der Gruppe dar, während $\Sigma(\theta_k)$ für die Kovarianzstruktur steht (vgl. Reinecke & Seddig, 2011: 419). Das Subskript k verweist auf die jeweilige latente Klasse. π_k bezeichnet hingegen die Wahrscheinlichkeit eines Falls zu der Gruppe k zu gehören, wobei sich diese Wahrscheinlichkeiten im Ergebnis auf 1 (bzw. 100%) aufaddieren müssen. ϕ_k repräsentiert hingegen die jeweilige Wahrscheinlichkeitsdichtefunktion von y_i . Die beobachteten Daten können somit als eine Funktion der Wahrscheinlichkeit der Klassenzugehörigkeit und einem Multigruppen-Wachstumsmodell repräsentiert werden (vgl. Ram & Grimm, 2009: 566). Während gruppenspezifische Unterschiede hinsichtlich der Mittelwertstruktur $\mu_k(\theta_k)$ zur Grundidee von Mischverteilungsmodellierungen gehören, gilt dies nicht in gleicher Weise für die jeweilige gruppeninterne Kovarianzmatrix $\Sigma(\theta_k)$. Die Parameter der Kovarianzmatrix (vorrangig die der Wachstumsparameter) können über die Gruppen hinweg sowohl frei geschätzt wie auch als invariant spezifiziert werden. Die hierbei getroffenen Annahmen für GMMs sind, dass innerhalb der latenten Klassen :

- konditionale Normalität der Verteilung der wiederholten Messungen vorherrscht;
- die Mittelwert- und Kovarianzstruktur korrekt spezifiziert ist;
- die Effekte evtl. externer Prädiktoren linear verlaufen;
- bei fehlenden Werten Missing at Random (MAR) vorliegt und
- die individuellen Fälle unabhängig sind (vgl. Bauer, 2007: 765 ff.).

2.1.1 Prinzip des EM-Algorithmus

Die dominanteste Methode zur Berechnung von Parametern im Rahmen von Mischverteilungsmodellen sind Maximum-Likelihood-Verfahren, insbesondere der Expectation-Maximization(EM)-Algorithmus (vgl. Duncan et al., 2002: 10). Auch Mplus verwendet zur

Bestimmung der Klassenzugehörigkeit einen EM-Algorithmus (vgl. Muthén & Shedden, 1999: 1 ff.). Ausgehend von Startwerten (bzw. den vorläufigen Parameterwerten) und Datenbasis wird hierbei bestimmt, mit welcher Wahrscheinlichkeit ein (Wachstumsverlaufs-)Fall einer latenten Gruppe zuzuordnen ist (E-Step). Nach Zuordnung der Fälle werden die vorläufigen Parameterwerte angepasst (M-Step) und ein neuer Iterationsschritt vollzogen bis die EM-Schätzung konvergiert. Da andere Startwerte potenziell zu anderen Endergebnissen und zu nicht konvergierenden Schätzungen führen können, gilt es bei der Berechnung von GMMs mehrere Startwertkombinationen zu verwenden.⁵ Eine besondere Sensitivität des EM-Algorithmus gegenüber Startwerten zeigt sich insbesondere im GMM Kontext. Da dem EM-Algorithmus ein asymptotisch erwartungstreuer Schätzer zugrunde liegt, benötigt er eine hohe Fallzahl um zu adäquaten Ergebnissen zu gelangen (vgl. Depaoli, 2013: 190). Für die Schätzung von GMMs ist die hohe Fallzahl im Idealfall mit einer klaren Trennung der vom GMM zu schätzenden Klassen verbunden, was jedoch in der Forschungspraxis nicht immer gegeben sein wird (vgl. Depaoli, 2013: 190).

Auch der gängige Anwendungskontext des EM-Verfahrens zur Bestimmung von Maximum-Likelihood-Schätzern bei fehlenden Werten (vgl. Dempster et al., 1977) kann konzeptionell mit der Bestimmung der Klassenzugehörigkeit im Rahmen von GMM-Schätzungen verbunden werden. Dazu muss man die latente Klassenzugehörigkeit lediglich als fehlenden Wert verstehen, welche für jeden individuellen Wachstumsverlauf der Untersuchungseinheiten fehlt und über die EM-Methodik bestimmt werden kann (vgl. Masyn, 2013: 554).

2.1.2 Modellkonstruktion

Für den Aufbau eines GMM werden aktuell in aller Regel mehrstufige modellbildende Analyseschritte vorgeschlagen (z. B. Ram & Grimm, 2009; Qureshi & Fang, 2010; Reinecke et al., 2015; Bauer & Curran, 2004; Wickrama, 2016). Auch wenn sich die Anzahl der modellbildenden Analyseschritte zwischen den einzelnen Quellen unterscheiden, lassen sich allgemeine Gemeinsamkeiten in den Empfehlungen ausmachen.

Abgesehen von analytischen Vorüberlegungen und deskriptiven Voranalysen einzelner zeitlicher Entwicklungsverläufe, beinhaltet der erste Schritt der Modellbildungsempfehlungen die Spezifikation eines herkömmlichen LGC-Modells als Baseline für die weiteren Analyseschritte, welche noch keine Unterscheidung nach evtl. vorhandenen Klassenunterschie-

⁵Das Statistiksoftwarepaket Mplus bietet hierzu einen eigenen Unterbefehl an, welcher die Angabe der Anzahl von unterschiedlichen randomisierten Startwertkombinationen ermöglicht (vgl. auch Muthén & Asparouhov, 2008: 152).

den macht. Ziel dieses ersten Analyseschrittes ist es (a) die beste funktionale Form des Wachstums für ein Einklassenmodell zu finden, (b) vorhandene Variation in den Wachstumsparametern zu identifizieren sowie (c) das Modell hinsichtlich der Passung auf die empirische Datenbasis zu bewerten. Schlechte Werte hinsichtlich der Modellpassung sowie Heterogenität in den Wachstumsparametern können hierbei Hinweise für die Existenz von relevanten Subpopulationen mit eigenen Wachstumsverläufen liefern (vgl. Wickrama, 2016: 202). Fehlende Varianz in den Wachstumsparametern oder eine sehr gute Passung des LGC-Modells wären hingegen Hinweise gegen die Existenz von latenten Wachstumsklassen hinsichtlich der untersuchten Größe.

Der zweite Schritt der Step-by-Step Anweisungen unterscheidet sich stärker zwischen den einzelnen Literaturempfehlungen. Dennoch stimmen die verschiedenen Autoren darin überein, dass die Schätzung einer Latent Class Growth Analysis (LCGA), wenn auch aufgrund jeweils anderer Überlegungen, anzuraten ist. LCGAs⁶ basieren auf den Arbeiten von Nagin et al. (vgl. Nagin & Tremblay, 1999; Muthén, 2004) und können als eine vereinfachte Form von GMMs verstanden werden (vgl. Muthén, 2004: 349 f.). So werden bei LCGA zwar im Gegensatz zu LGC-Modellen mehrere Subpopulationen zugelassen, aber die Variabilität der Wachstumsparameter innerhalb einer identifizierten Subpopulation ist auf eine Nullmatrix fixiert (vgl. Bauer & Curran, 2004: 14). Folglich werden zwar für jede latente Gruppe unterschiedliche Mean-Werte für Steigung(en) und Intercept geschätzt, aber für die einzelnen Fälle innerhalb der Wachstumsklasse werden keine Abweichungen für diese klassenspezifischen Werte zugelassen (bzw. als Zufallsfehler verstanden). Eine Folge dieser Spezifikation ist, dass LCGA die Variabilität der Daten (sowohl inhaltlich substantielle wie auch zufällige Abweichungen) dadurch verarbeiten, dass sie mehr latente Klassen identifizieren als aus einer inhaltlich-substanziellen Perspektive zu vertreten wäre. Simulationsstudien weisen deutlich darauf hin, dass Verwendung von LCGA im Durchschnitt zur Schätzung von zu vielen latenten Klassen führen und daher von der alleinigen Verwendung von LCGA abzuraten ist (vgl. Diallo et al., 2016: 20; Bauer & Curran, 2003: 357 ff.). Allerdings kann die Verwendung im zweiten Analyseschritt für GMMs insofern hilfreich sein, als dass eine Obergrenze für die potenzielle Anzahl von latenten Klassen festgelegt werden kann (vgl. Muthén, 2004: 350; Reinecke et al., 2015: 72). Die Anzahl der latenten Klassen der LCGA-Modellierung dient somit als obere (Baseline-)Vergleichsgröße

⁶Der Begriff Latent Class Growth Analysis wurde allerdings von Muthén (2004) geprägt, um die stärkere Verwandtheit mit der latenten Klassenanalyse als mit der Growth Modellierung herauszustellen. Nagin & Tremblay (1999) selbst sprechen von Group-Based Trajectory Modeling.

für Growth Mixture Modellierung wie bereits die funktionale Wachstumsform des LGC-Modells im ersten Modellierungsschritt.

Der dritte Schritt beinhaltet schließlich die Schätzung des eigentlichen Growth Mixture Modells, welches nicht nur Gruppenunterschiede hinsichtlich der Wachstumsverlaufsmuster zulässt, sondern auch Variabilität der Intercept und Slope Parameter innerhalb der latenten Gruppen. Aber auch die Schätzung eines GMM gestaltet sich in der Ausführung keineswegs als trivial, da trotz der vorangegangenen Analyseschritte noch einige Entscheidungen hinsichtlich der tatsächlichen Modellspezifikation und Schätzung zu treffen sind. So wird, beispielsweise, vom Statistikpaket Mplus die Kovarianzmatrix für die verschiedenen latenten Gruppen als Defaulteinstellung invariant spezifiziert bzw. geschätzt. Zwar reduziert diese Vereinfachung die Rechenlaufzeit und die Chancen der Modellkonvergenz, aber Simulationsstudien haben anhand ähnlicher Restriktionen der Kovarianzmatrizen bereits zeigen können, dass dies im Durchschnitt zu einer Extraktion zu vielen latenten Klassen und zu Verzerrungen hinsichtlich der Parameterschätzwerte führt (vgl. Wickrama, 2016: 205)(siehe auch: Enders & Tofghi, 2008; Bauer & Curran, 2004; Magidson & Vermunt, 2004). Folglich sind auch im dritten Analyseschritt Abwägungen hinsichtlich Komplexität, Konvergenz und Modellpassung zu treffen. Aber auch nach Identifikation der angemessenen Modellspezifikation muss noch die Frage nach der korrekten Anzahl an latenten Klassen beantwortet werden.⁷

2.1.3 Identifikation der Anzahl latenter Klassen

Da die Anzahl der latenten Klassen im Rahmen einer GMM-Analyse nicht im vornherein festgelegt ist, gilt es zunächst Kriterien zu finden, welche die korrekte Anzahl an latenten Klassen identifizieren. Die bisher im Einsatz befindlichen Analysealgorithmen sind noch nicht dazu im Stande dem/der ForscherIn, auf Grundlage einer einfachen Rechenoperation, die den Daten angemessene Anzahl an latenten (Wachstumskurven-)Klassen auszugeben. Vielmehr ist es im Rahmen einer GMM Analyse nötig, mehrere Schätzungen auf Grundlage einer variierten Anzahl an unterstellten latenten Klassen durchzuführen und im Anschluss daran anhand von Kriterien und statistischer Tests zu entscheiden, welche Anzahl an latenten Klassen die beste Lösung darstellt.

⁷Die Ausführungen in diesem Abschnitt wurden in abgewandelter Form bereits bei Krause et al. (2019) veröffentlicht (vgl. Krause et al., 2019: 393 ff.).

Da ein Mixture Modell mit k latenten Klassen nicht⁸ in einem ansonsten äquivalenten Modell mit $k+1$ Klassen genestet ist, kann man für den Modellvergleich keine klassischen χ^2 -Differenztest verwenden (vgl. Reinecke & Seddig, 2011: 420). Da GMMs zudem für jede latente Klasse eine eigene Kovarianzmatrix ausgeben, lassen sich χ^2 -basierte Indizes wie der CFI (comparative fit index) oder der RMSEA (root mean square error of approximation) nicht sinnvoll berechnen (vgl. McNeish & Harring, 2017: 228), wodurch auch die Identifikation der richtigen Klassenanzahl nicht darauf beruhen kann. Für den Modellvergleich und die Entscheidung über die korrekte Anzahl an latenten Klassen müssen somit andere Entscheidungsgrößen verwendet werden. Die Kriterien zur Identifikation der korrekten Anzahl an latenten Klassen lassen sich in drei Kategorien aufgliedern:

- (a) Information Criteria Statistics,
- (b) Goodness of Fit Test bzw. Entropy Values,
- (c) Nested Model Tests bzw. Likelihood Ratio Tests

(vgl. Wickrama, 2016: 213; Tofghi & Enders, 2008: 319 ff.; Peugh & Fan, 2012: 206).

(ad a) Information Criteria Maßen (IC-Maße) werden generell anhand der Werte von Devianz (bzw. $-2 * \text{Log-Likelihood (LL)}$), der Anzahl an geschätzten Modellparametern (p) und der Fallzahl (N) berechnet (vgl. Peugh & Fan, 2012: 206 f.). Somit wird bei der Berechnung der IC-Werte anhand der Devianz sowohl nach der Modellkomplexität (gemessen durch die Parameteranzahl) als auch nach der jeweiligen Fallzahl korrigiert (vgl. Wickrama, 2016: 213 f.). Mit der Ausnahme des Akaikes Information Criteria (AIC) haben alle IC-Maße die statistische Eigenschaft der Konsistenz, welche dazu führt, dass bei steigender Fallzahl häufiger das korrekte Modell in der Grundgesamtheit identifiziert werden kann (vgl. Peugh & Fan, 2012: 207). Allgemein deuten geringere Werte der IC-Maße auf eine bessere Passung des Modells (bzw. der Anzahl der latenten Klassen eines Modells) hin, auch wenn hierfür keine klaren Grenzwerte angegeben werden können. Zu den wichtigsten und am häufigsten eingesetzten Informationskriterien zählen der AIC und das Bayesian Information Criterion (BIC) sowie diverse Modifikationen dieser Maßzahlen, wie das Sample-Size Adjusted BIC (aBIC/SSABIC) (vgl. Sclove, 1987) oder das Consistent AIC (vgl. Bozdogan, 1987). Die Formeln dieser zentralen IC-Maße sind hierbei wie folgend definiert:

⁸Zumindest in keiner praktisch relevanten Weise. Vgl. auch Fußnote 9 und 10 und den dazugehörigen Textbezug.

$$AIC = -2LL + 2 * p \quad (4)$$

$$CAIC = -2LL + p * (\log(n) + 1) \quad (5)$$

$$BIC = -2LL + p * \log(n) \quad (6)$$

$$aBIC/SSABIC = -2LL + p * \log\left(\frac{n+2}{24}\right) \quad (7)$$

(vgl. Nylund et al., 2007: 545). Wobei LL für den logarithmierten Likelihoodwert, p für die Anzahl der Modellparameter und n für die Fallzahl steht. Da kleinere Werte eine bessere Passung widerspiegeln, bestrafen diese IC-Maße hohe Modellkomplexität (tlw. im Verhältnis zur Fallzahl).

(ad b) Goodness of Fit Test bzw. Entropy Werte können als ergänzendes Kriterium der Modellselektion verstanden werden. Sie basieren auf der jeweiligen A-posteriori-Wahrscheinlichkeit der Klassenzuordnung der einzelnen empirischen Wachstumsfälle (vgl. Qureshi & Fang, 2010: 225). Entropie ist als ein standardisiertes Maß für die Genauigkeit bzw. Unschärfe der Klassenzuordnung im Rahmen von GMMs zu verstehen (vgl. Diallo et al., 2016: 6). Die möglichen Werte der Entropie bewegen sich im Bereich von 0 bis 1, wobei 0 implizieren würde, dass keinerlei Sicherheit bei der Fallzuordnung zu latenten Wachstumsklassen besteht, wohingegen ein Wert von 1 bedeuten würde, dass für die Fallzuordnung völlige Sicherheit vorherrscht und eine perfekte Abgrenzung/Separation der latenten Klassen gegeben ist. Die Berechnung der Entropie erfolgt nach der Formel

$$E_k = 1 - \frac{\sum_i \sum_k (-p_{ik} * \ln p_{ik})}{N - \ln k} \quad (8)$$

wobei p_{ik} für die geschätzte Wahrscheinlichkeit des Falles i , in der Gruppe k zu sein steht (vgl. Reinecke et al., 2015: 77). Für Modelle mit ähnlicher Modellpassung kann folglich empfohlen werden, dass höhere Entropie Werte bevorzugt werden sollten, da dies auch mit dem praktischen Ziel der Zuordnung von Fällen in latente Klassen zusammenfällt (vgl. Ram & Grimm, 2009: 571). Allerdings kann die Entropie nicht als ein absolutes Kriterium der Klassenextraktion verwendet werden. Da die Entropie eine Funktion der Anzahl von latenten Klassen, der Variabilität innerhalb dieser Klassen sowie der Unterschiede zwischen den Klassen darstellt, würde dies im Ergebnis zu einer deutlichen Überextraktion

führen. Würde man, beispielsweise, genau so viele Klassen wie empirische Fälle modellieren, würde dies ebenfalls zu einem perfekten Wert der Entropie führen, da jeder Fall seine eigene Klasse bekommen würde und damit perfekt zugeordnet wäre (vgl. Hancock & Samuelsen, 2008: 65). Bisherige Empfehlungen gehen bei Grenzwerten von 0.40 von einer niedrigen, bei 0.60 von einer mittleren und bei 0.80 von einer hohen Klassenseparation aus (vgl. Wickrama, 2016: 214). Neben dem rohen Entropie-Wert existieren auch Indizes, welche neben der Devianz zusätzlich nach dem Grad der Entropie kontrollieren (Normalized Entropy Criterion (NEC)(vgl. Celeux & Soromenho, 1996) und Classification Likelihood Criterion (CLC) (vgl. Biernacki & Govaert, 1997) und somit GMM Modelle bevorzugen, welche eine klare Wachstumsklassenzuordnung aufweisen (vgl. Peugh & Fan, 2012: 207). Darüber hinaus kombiniert das Integrated-Completed Likelihood Criterion (ICL-BIC) (vgl. Biernacki et al., 2000) die entropie-basierten Indizes mit IC-Maßen indem er zusätzlich Modellkomplexität bestraft:

$$NEC = \frac{Entropie}{LL_k - LL_{k=1}} \quad (9)$$

$$CLC = -2LL + 2 * Entropie \quad (10)$$

$$ICLBIC = -2LL + p * \ln(N) + 2 * Entropie \quad (11)$$

(vgl. Peugh & Fan, 2012: 208).

(ad c) Likelihood-Ratio basierte Testverfahren setzen an der Idee des klassischen χ^2 -Differenztest an und bieten ein formalisiertes Testverfahren für den Vergleich von Modellen mit einer unterschiedlichen Anzahl an latenten Klassen. Der LRT betrachtet die Log-Likelihood-Differenz genesteter Modelle als Teststatistik, die unter bestimmten Bedingungen asymptotisch einer χ^2 -Verteilung folgt, deren Freiheitsgrade gleich der Differenz in der Anzahl der Parameter der beiden genesteten Modelle sind. Modelle mit k-1 und k vielen latenten Klassen können zwar formal ineinander genestet sein⁹, aber der Vergleich von latenten Klassenmodellen entspricht dabei nicht bestimmten Regelmäßigkeitsannahmen¹⁰

⁹Wenn der Anteil einer Klasse auf Null fixiert wird und bei Invarianzannahme hinsichtlich klassenspezifischer Parametern in zwei Klassen (vgl. Tekle et al., 2016: 210).

¹⁰Dies betrifft hier zwei mögliche Annahmen: (1) Die Parameterrestriktionen unter dem Nullmodell müssen ein innerer Punkt (und damit keine Grenze) des zulässigen Parameterbereichs sein; (2) alternative Spezifikationen zur Nestung von Modellen mit K-Klassen und K-1-Klassen, welche klassenspezifische Parameter im k- und k-1-Modell gleichsetzen würde, verletzt eine andere Regelmäßigkeitsbedingung, nämlich, dass die Informationsmatrix nicht singular sein darf (vgl. Tekle et al., 2016: 210)(siehe auch: McLachlan & Peel, 2000: 185 ff.; Aitkin & Rubin, 1985: 70).

für einen traditionellen LR-Differenztest mit einer χ^2 -Verteilung (vgl. Tekle et al., 2016: 210). Daher muss zum Modellvergleich auf alternative Testverfahren zurückgegriffen werden.

Der (Adjusted) Lo-Mendell-Rubin Test (ALMR) (vgl. Lo, 2001) und der Bootstrapped Parametric Likelihood Ratio Test (BLRT) (vgl. McLachlan & Peel, 2000) bieten Lösungen an, welche wiederum einen Signifikanztest zwischen Modellen mit einer unterschiedlichen Anzahl von latenten Klassen zulassen. Nach der Testlogik wird hierbei so verfahren, dass ein Modell mit $k-1$ Klassen einem Modell mit k Klassen gegenübergestellt wird, wobei die jeweils zu verwerfende Nullhypothese ist, dass das Modell mit weniger Klassen mindestens genauso gut auf die Daten passt wie das komplexere Modell. Während der ALMR mit einer Approximation der Referenzverteilung für den Likelihood Ratio Test arbeitet, basiert der BLRT auf einer empirischen Referenzverteilung (vgl. Tofghi & Enders, 2008: 320 f.). Beide Verfahren gehen hierbei von multivariater Normalverteilung aus, wobei eine Verletzung dieser Annahme in der Tendenz zu einer Überextraktion an latenten Klassen führen würde (vgl. Peugh & Fan, 2012: 208).

Neben formalen Selektionskriterien der Modellselektion werden allerdings auch weitere Entscheidungsheuristiken für die Forschungspraxis empfohlen, welche die Interpretierbarkeit, die Übereinstimmung mit dem theoretisch/ analytischen Rahmen sowie die Größe der latenten Klassen betreffen. So sollte bei sehr ähnlichen Fit-Werten im allgemeinen das sparsamere bzw. weniger komplexe Modell bevorzugt werden. Auch die Interpretierbarkeit der Klassen sowie die Übereinstimmung mit theoretischen und/oder analytischen Vorüberlegungen sollte die Selektion der Klassenanzahl informieren. Darüber hinaus sollte in der Regel davon abgesehen werden, dass sehr kleine Klassen, welche weniger als 5% der Fälle oder unter 25 Beobachtungen enthalten als substantiell bedeutsam zu interpretieren, wenn hierfür nicht sehr gute Gründe anzuführen sind (vgl. Wickrama, 2016: 216 f.).

Diese vermeintliche Subjektivität bei der Modellselektion – bzw. die Unzufriedenheit darüber – durch weiche Kriterien oder die Auswahl des jeweils passenden Enumeration Indizes inspiriert auch die Entwicklung von alternativen Verfahren der Modellselektion, welche sich in das bisherige Schema der Entscheidungskriterien noch nicht einordnen lassen. Aktuelle Arbeiten im Bereich der Mixture Models schlagen z. B. alternative Modellselektionskriterien vor, welche auf der Idee der Kreuzvalidierung beruhen. Grimm et al. (2017) präsentieren die sogenannte k -Fold Cross-Validation Technik, welche die Stichprobe zunächst in k -viele nicht-überschneidende Teil untergliedert (vgl. Grimm

et al., 2017: 4). Die Parameter des Analysemodells werden dann anhand der $(k - 1)$ -Teilstichproben (auch: Trainings- oder Kalibrierungs-Sample) berechnet. Das Analysemodell wird im nächsten Schritt dann mit den errechneten Parameterwerten auf die verbleibende k -Stichprobe (auch: Test-, Validierungs- oder Holdout-Sample) angewendet um ein (vorher zu wählendes) Fitmaß (z. B. $-2LL$) zu errechnen. Dieser Prozess wird für jede Klassenanzahl k -Mal wiederholt, wodurch jede Kombinationsmöglichkeit der Teilstichproben durchdekliniert wird (vgl. Grimm et al., 2017: 4). Die daraus generierten Verteilungen der Fitmaße können dann entlang von verschiedenen Entscheidungsheuristiken¹¹ genutzt werden, um die optimale Klassenlösung zu identifizieren. Ob durch diese Technik bessere, verlässlichere oder weniger idiosynkratische Ergebnisse als durch die zuvor diskutierten Standardverfahren generiert werden können, bleibt, wie die Autoren auch anerkennen, noch offen, da eine systematische Evaluation der k -Fold Cross-Validation Technik in weiten Teilen noch aussteht. Erste Simulationsergebnisse im Kontext von GMMs weisen allerdings darauf hin, dass die k -Fold Cross-Validation Technik nur bei hoher Klassendistanz zuverlässig funktioniert und insgesamt mit geringen korrekten Identifikationsraten zu kämpfen hat (vgl. He & Fan, 2018: 12). Mangels nachweisbarer Performanz und fehlender Softwareimplementation kann aktuell nur von dem Einsatz der k -Fold Cross-Validation Technik abgeraten werden.

2.2 GMMs als MIGMMs für die Sozialforschung

Als Multiple Indicator Growth Mixture Models (MIGMMs) werden Modelle bezeichnet, welche für die einzelnen Messzeitpunkte nicht einen manifesten Indikator oder einen Composite Score wie beim klassischen GMM, sondern ein latentes Konstrukt verwenden (vgl. Abb. 4). MIGMMs (oft auch als Second Order Growth Mixture Model bezeichnet) lassen sich aktuell in der empirischen Anwendung noch sehr selten finden. Bei den wenigen Ausnahmen hiervon handelt es sich um Paper, welche vorrangig die Anwendung demonstrieren, aber dafür weniger inhaltlich substanziell motiviert vorgehen (z. B. Ram & Grimm, 2009; Hancock & Samuelsen, 2008).¹² Eine systematische Untersuchung der Performanz von MIGMMs durch Simulationsstudien findet sich bis vor Kurzem nur bei Zhang (2016) und ist somit in der Gesamtbetrachtung noch mit vielen offenen Fragen behaftet (siehe

¹¹Grimm et al. (2017) diskutieren hierzu vier verschiedene Entscheidungsverfahren, wobei auch weitere Alternativen denkbar wären.

¹²Für Beispiele von Ausnahmen zur raren empirischen Nutzung siehe Kapitel 3.1.

auch Ram & Grimm, 2009: 139).¹³ Diese spärliche Nutzung in der empirischen Praxis ist insofern bemerkenswert, als dass die Nutzung von latenten Konstrukten zu den einzelnen Messzeitpunkten mit einer Reihe von Vorteilen verbunden ist, welche auch den Mixture-Part von GM-Modellen positiv beeinflussen können.

Durch die Modellierung von latenten Messmodellen kann sowohl nach unsystematischen als auch nach zeitspezifischen Messfehlern der Indikatoren kontrolliert werden, wodurch (unter anderem) die Reliabilität der Messungen erhöht wird (vgl. Urban & Mayerl, 2014: 16; Urban, 2004: 6). Für (MI)GMM ist dies insofern relevant, als dass hierdurch unechte (also nur messfehlerbedingte) Varianz und Parameterverzerrungen reduziert werden können. Infolgedessen können latente Wachstumsgruppen und deren Verlaufsmuster genauer bestimmt werden und damit Fehlinterpretationen durch Schätzung von zu vielen oder zu wenigen Gruppen vorgebeugt werden. MIGMMs verwenden daher sozusagen nur den reliablen Varianzanteil der empirischen Daten.

Ein weiterer zentraler Vorteil von MIGMMs liegt in der Überprüfbarkeit und evtl. Aufhebung von Messinvarianzannahmen. Ein-Indikator-Messungen (auch wenn es sich hierbei um Composite Scores handelt) gehen implizit davon aus, dass die Messung zu jedem Erhebungszeitpunkt identisch erfolgt (siehe auch Urban, 2004: 6). Im Kontext von GMMs gehen diese Messinvarianzannahmen sogar noch einen Schritt weiter. Es wird nicht nur angenommen, dass die Messungen zeitinvariant erfolgen, sondern zusätzlich zwischen den identifizierten latenten Wachstumsverlaufgruppen keine Unterschiede aufweisen. Durch die Modellierung von MIGMMs können sowohl zeit- wie auch gruppenspezifische (partielle) Faktorinvarianzen überprüft und zugelassen werden. Auf Grundlage dieser Möglichkeit können somit tatsächliche Entwicklungsverläufe sowie deren gruppenspezifischen Unterschiede untersucht werden, welche nicht nur durch eine Varianz in der Bedeutung des Messinstruments erzeugt werden.

Eine Besonderheit ergibt sich bei der Schätzung von GMM hinsichtlich der empirischen Verteilung von wiederholten Messungen einer latenten Gruppe. Sollte für diese empirischen Verteilungen keine (konditionale) Normalverteilung vorherrschen, führt dies im Ergebnis zu einer Überextraktion von artifiziellen Wachstumsklassen (siehe Abschnitt 3.2.3). Dies ist besonders problematisch, wenn in der tatsächlichen Population, aus welcher die empirische Verteilung stammt, eigentlich keine Abweichungen von der Normalverteilung

¹³Eine aktuelle Simulationsstudie zu Gruppenunterschieden bei MIGMMs von Kim & Wang (2017) wird später ausführlicher diskutiert.

vorherrschen. So besteht bei der Messung durch einen manifesten Indikator die erhöhte Möglichkeit, dass die Nicht-Normalverteilung lediglich durch Decken- und Bodeneffekte oder andere Anomalien der Messung bedingt ist (vgl. Grimm & Ram, 2009: 125). Diese möglichen Gründe für nicht-normalverteilte Daten können durch mehrfache Messungen und der Spezifikation als latentes Messkonstrukt behoben werden. Indem hierbei Messfehler kontrolliert werden, wird ein genaueres Abbild der tatsächlichen Verteilung in der Population wiedergegeben. Dadurch werden bei einem GM-Modell evtl. weniger falsche Extraklassen extrahiert, welche den Bedingungen in der Population nicht entsprechen. Die Vorteile der Modellierung als Second Order (bzw. Multiple Indikator) Konstrukt treffen also sowohl auf die generelle Messgenauigkeit zu, als auch auf den genuinen Zweck der GMM Spezifizierung: die Identifikation von latenten Klassen als ein Ausdruck von tatsächlicher inhaltlicher Heterogenität in der Population.

3 Bisherige Ergebnisse

Dieser Abschnitt soll zunächst bisherige empirische Anwendungsfelder von GMMs und MIGMMs/SOGMMs benennen um einen Überblick zur aktuellen Anwendung in der empirischen Forschung zu gewähren und einen Ausblick auf zukünftige Entwicklungen zu ermöglichen. In einem zweiten Schritt soll darauf eingegangen werden, welche Ergebnisse sich anhand bisheriger Arbeiten hinsichtlich der Performanz von GMMs und MIGMMs ziehen lassen. Dabei werden sowohl Erfahrungen aus der empirischen Anwendung als auch Ergebnisse aus nicht-empirischen Simulationsstudien berücksichtigt.

3.1 Überblick zu empirischen Anwendungsbereichen von (MI)GMMs

In der letzten Dekade hat die Growth Mixture Modellierung zunehmend an Popularität gewonnen (vgl. Diallo et al., 2016: 1). Die Anwendung von GMM Modellen bietet sich allerdings vorrangig in Analysesettings an, für welche angenommen werden kann, dass ein zeitlicher Entwicklungsverlauf vorherrscht. Anwendungsbereiche, welche in der abhängigen Größe eine zeitliche Konstanz aufweisen, eignen sich hingegen wenig für GMMs. Neben einem Entwicklungsverlauf benötigt der Einsatz von GMMs allerdings auch die Annahme über Unterschiede hinsichtlich dieser Entwicklungsverläufe. Nur auf der Grundlage von intraindividuellem Wandel und interindividuellen Unterschieden hinsichtlich dieses

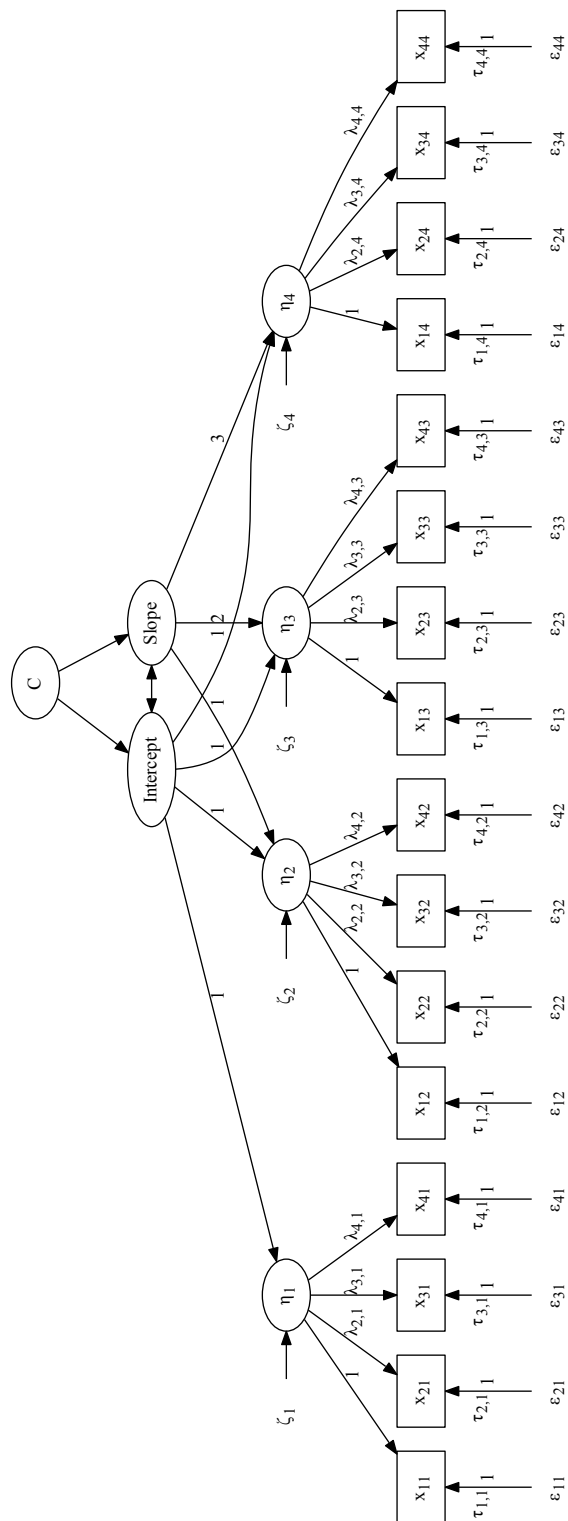


Abbildung 4: Grundstruktur eines MIGMM-Modells mit einer linearen Steigung.

Diachrone Fehlerkorrelationen und die vollständige Mittelwertstruktur sind aus Übersichtlichkeitsgründen nicht abgebildet. Eigene Darstellung.

Wandels bietet sich der Einsatz von GMM an. Dieser Wandel kann durch die Growth Mixture Modellierung identifiziert und quantifiziert werden. Darüber hinaus erlaubt die GM-Modellierung auch die gruppenspezifische Analyse von Variablen, welche die latente Gruppenzugehörigkeit beeinflussen oder den Wachstumsverlauf innerhalb der latenten Gruppe determinieren. In Übereinstimmung mit diesen Überlegungen haben GMMs bereits in empirischen Studien im medizinischen Bereich, in Untersuchungen zu Entwicklung aggressiven Verhaltens von Kindern und Jugendlichen, in der Bildungsforschung und allgemein im Kontext der Devianzforschung Anwendung gefunden (vgl. Muthén & Asparouhov, 2008: 143 ff.; Hancock & Samuelsen, 2008: 5).

Im Schnittbereich zwischen Medizin und Sozialforschung finden sich Anwendungen, welche sich mit Alkoholmissbrauch (z. B. Greenbaum et al., 2005), mit dem Verlauf von Symptomen von Depression sowie deren Prädiktoren und Möglichkeiten der Behandlung (z. B. Stoolmiller et al., 2005; Leuchter et al., 2002) oder auch mit Medikamenten-Wirkungen (z. B. Muthén et al., 2011) auseinandersetzen.

Im Bereich der Sozialwissenschaften nimmt die Devianz- und Suchtforschung hinsichtlich des Einsatzes von GMMs eine besondere Rolle ein. So sind in diesen Bereichen sowohl Entwicklungsverläufe identifizierbar, da sich die Untersuchungen in der Regel auf junge Kohorten beziehen, als auch Unterschiede bzgl. dieser Entwicklungsverläufe zwischen einzelnen Individuen. So finden sich Arbeiten, welche sich aus der Modellierung ergebenden Vorteile in der Interpretation von kriminellem Verhalten hervorheben (z. B. Kreuter, Muthen, 2008a), Untersuchungen, welche tatsächliche Entwicklungsverläufe und deren Gruppierung beleuchten (z. B. Reinecke, 2014; Reinecke et al., 2015), aber auch Analysen, welche Prädiktoren der Zugehörigkeit zu einer latenten Gruppe sowie deren Konsequenzen besondere Beachtung schenken (z. B. Petras & Masyn, 2010; Nagin & Tremblay, 1999). Auch für den Bereich Suchtforschung finden sich Arbeiten im Rahmen der GM-Modellierung. So untersuchen Li et al., beispielsweise, gruppenspezifische Unterschiede hinsichtlich des Effekts von Peer-Gruppen auf den Alkoholkonsum bei adoleszenten Personen (vgl. Li et al., 2002), während Wuk/Witkiewitz mit der Behandlung von Alkoholabhängigkeit im Kontext von sozialen Unterstützungsnetzwerken befassen (vgl. Wu & Witkiewitz, 2008). Windle und Wiesner untersuchen Muster des Marihuanakonsums von jungen Erwachsenen und identifizieren dabei, neben den Gruppenunterschieden im Verlauf, sowohl Prädiktoren für die Gruppenzugehörigkeit als auch Folgen im Sinn von späteren psychischen Störungen (vgl. Windle & Wiesner, 2004).

Im Bereich der Bildungsforschung wird mit dem Analyseinstrument GMM ein besonderes Augenmerk auf interindividuelle Unterschiede in Entwicklungsverläufen gelegt. So finden sich hier Untersuchungen zur Lesefähigkeit von Kindern (z. B. Douglas, Liu, 2009; Boscardin et al., 2008) sowie mögliche Ansatzpunkte zur Reduktion von Gruppenunterschieden (vgl. Parrila et al., 2005). Weitere Anwendungen beschäftigen sich mit dem Risiko des Schulabbruchs bei US-Schülern und dessen Abhängigkeit von Leistungen im Fach Mathematik sowie mit schulspezifischen Unterschieden hinsichtlich dieses Effekts (vgl. Muthén, 2004).

Bei allen bisher angeführten Studien handelt es sich allerdings um GMMs mit einer einzelnen manifesten Messvariable zu einem Erhebungszeitpunkt, ohne Verwendung eines latenten Messmodells. Bestenfalls wurde auf der Ebene des einzelnen Messzeitpunktes ein Composite Score gebildet. Trotz offensichtlicher Vorteile sind MIGMMs bzw. Second Order Modelle im Rahmen von empirischen Anwendungen von Growth Mixture Modellierungen immer noch die Ausnahme. Bisherige Anwendungen haben oft noch einen illustrativen Charakter und legen den Fokus auf das Modellierungsverfahren selbst und weniger auf die inhaltlich-substantiellen Aspekte (z. B. Grimm & Ram, 2009; Hancock & Samuelson, 2008). Aber es lassen sich auch einzelne Ausnahmen von dieser Regel finden. Callina et al. (2014) untersuchen z. B. Hoffnung und Vertrauen als MIGMMs bei Jugendlichen und deren Auswirkung auf späteres soziales Engagement. Dabei können sie vier distinkte Entwicklungsverlaufsklassen identifizieren (vgl. Callina et al., 2014: 877 f.). Kirves et al. (2014) beschäftigen sich hingegen mit der wahrgenommenen Beschäftigungsfähigkeit auf Grundlage eines MIGMMs und deren Auswirkungen auf die Arbeitszufriedenheit, wobei sie ebenfalls bis zu vier Wachstumsentwicklungsklassen identifizieren können. Auch Krause et al. (2019) modellieren anhand einer Jugendstichprobe eine MIGMM. Der inhaltliche Fokus liegt dabei auf der Entwicklung von sozialer Exklusion, Prädiktoren des Entwicklungsverlaufs von Exklusion und späteren Auswirkungen auf pro-soziale Verhaltensmuster. Dabei können zwei distinkte Entwicklungsverläufe ausgemacht werden, wobei sich aus methodischer Perspektive zeigt, dass MIGMMs auch bei vergleichsweise restriktiven Modellannahmen, in der empirischen Praxis, hinsichtlich der Modellkonvergenz Probleme aufweisen können (vgl. Krause et al., 2019: 400 ff.).

3.2 Überblick zu bisherigen Ergebnissen zur Performanz

Die bisherigen Ergebnisse zur Leistungsfähigkeit von GMMs und deren Probleme lassen sich in mehrere Teilbereiche aufschlüsseln. Dennoch können diese einzelnen Themenbereiche nicht gänzlich isoliert betrachtet werden, da Abhängigkeiten im Sinne von Wechselwirkungen und Moderationseffekten zwischen einzelnen Performanzaspekten sowie der Datenlage (oder der Simulationskonfiguration) bestehen. Daher müssen bisherige Ergebnisse auch immer vor dem Hintergrund der jeweiligen Prüf- bzw. Datenbedingungen interpretiert werden.

3.2.1 Maße der Klassenidentifikation

Wie bereits aufgeführt, liegt eine Vielzahl an unterschiedlichen Maßen vor, welche die Identifikation der korrekten Anzahl an latenten Klassen anleiten kann. Im Idealfall sollte sich anhand von Simulationsstudien ermitteln lassen, welche Maße die höchste Trefferquote bei der Identifikation von latenten Klassen aufweisen, um die Auswahl auf eine geeignete Menge zu beschränken. Nicht zuletzt aus diesem Grund beschäftigen sich zahlreiche Arbeiten mit diesem Aspekt. Trotz (oder auch gerade wegen) mehrerer Simulationsstudien, welche sich mit der Suche nach dem geeignetsten Maß beschäftigen, ist noch keine Einigkeit hergestellt. Allerdings lässt sich hinsichtlich des Vergleichs von IC-Maßen und LRT-Tests eine Tendenz identifizieren. Der BIC/aBIC sowie der BLRT schneiden unter vielen Simulationsbedingungen im Verhältnis zu den anderen Maßzahlen relativ gut ab. Nylund et al. (2007) machen anhand ihrer Simulationsstudien den BLRT als den konsistentesten Indikator für die korrekte Klassenzahl aus (vgl. Nylund et al., 2007: 564). Peugh & Fan (2012) kommen zu einem ähnlichen Ergebnis, wobei der BLRT insbesondere bei komplexen Wachstumsverläufen gut abschneidet (vgl. Peugh & Fan, 2012: 223). Liu & Hancock (2014) können für den BLRT (und eine Variante des LMR) ab einer Fallzahl über 700 perfekte Identifikationsraten ausmachen (vgl. Liu & Hancock, 2014: 570). Einen Kontrapunkt hierzu setzt allerdings Diallo et al. (2016). Anhand verschiedener Simulationskonfigurationen zeigte sich, dass der BLRT (wie auch der LMR und ALMR) durchweg geringere Trefferquote bei der Klassenidentifikation aufzeigten wie die untersuchten IC-Maße (vgl. Diallo et al., 2016: 20). Auch aktuelle Simulationsstudien bzgl. der Berücksichtigung von Kovariaten bei der Klassenidentifikation mehrten Zweifel an der noch anfangs unterstellten (sehr guten) Performanz des BLRT (vgl. Diallo et al., 2017:

43). Ein zusätzliches Problem des BLRT ergibt sich aus seiner Anfälligkeit hinsichtlich nicht normalverteilten Daten und Modellkomplexität. Daran sind, vor allem, fehlende Modellkonvergenz sowie suboptimale lokale Lösungen schuld (vgl. Wickrama, 2016: 215). Falls Modellschätzungen zu keiner (konvergierenden) Lösung kommen, können auch keine Vergleiche zwischen dem Modell mit $k-1$ und k vielen Klassen berechnet werden. Auch lokale Lösungen führen zu einer Verzerrung des LL-Werts und damit in Folge zu einem nicht angemessenen p -Wert des BLRT.

Hinsichtlich der IC-Maße scheint der aBIC die zuverlässigsten Ergebnisse zu liefern. Bei Nylund et al. (2007) und Tofghi & Enders (2008) zeigt er sich jeweils als zuverlässigster (IC-)Indikator der Klassenzahl. Auch Henson et al. (2007) identifizieren den aBIC als Maßzahl mit der höchsten Trefferquote hinsichtlich der Klassenanzahl. Der aBIC wurde hierbei Varianten des LMR sowie entropiekorrigierten Maßzahlen gegenübergestellt. Dabei zeigte sich, dass der aBIC konsistent die besten Ergebnisse liefert, wobei allerdings alle Maße bei einer zu geringen Fallzahl (hier ab 500) zu ungenau wurden (vgl. Henson et al., 2007: 221 ff.). Diallo et al. (2016) kommen hingegen zu dem Schluss, dass keines der untersuchten IC-Maße (CAIC, BIC, SCAIC, aBIC) zu konsistent besseren Ergebnissen kommt. Da die IC-Maße jeweils in unterschiedlichen Daten- und Spezifikationssituationen Vor- und Nachteile aufweisen, sehen sie diese weitestgehend als komplementäre Maße an (vgl. Diallo et al., 2016: 20). Eine klare Ausnahme von diesem dispersen Bild stellt der einfache AIC dar. In keiner gesichteten Untersuchung lieferte er adäquate Ergebnisse und kann somit als weitestgehend nutzlos angesehen werden. Alternative Verfahren wie die k -fold Cross-Validation wurden noch nicht umfassend evaluiert, wobei erste Ergebnisse eine sehr unbefriedigende Performanz andeuten (vgl. He & Fan, 2018: 10 ff.). Da noch kein Index oder Test identifizierbar ist, welcher in jeder möglichen Konfiguration der Daten die besten Ergebnisse liefert, sollten weiterhin mehrere Maße Anwendung finden. Im Idealfall werden dabei die Schwächen und Stärken der Maße berücksichtigt, indem die Datenlage daraufhin untersucht wird und die jeweils angemessensten Indikatoren und Tests Anwendung finden.

3.2.2 Klassenseparation und Klassenanteil

Die Klassenseparation im Rahmen einer GMM Analyse ist sowohl über die Unterschiede hinsichtlich der latenten Mittelwerte der Wachstumsparameter bestimmt als auch über die Größe der Wachstumsparametervarianz sowie deren evtl. Überschneidungen zwischen den

Klassen. Die Unterschiede zwischen den Klassen hängen also sowohl von den durchschnittlichen Wachstumsverlaufsmustern ab, als auch von der Variabilität innerhalb der Klassen. Besondere Probleme hinsichtlich der Klassenzuordnung sind unter Bedingungen geringer Klassenseparation zu erwarten. Unterscheiden sich die Wachstumsverläufe nicht deutlich, erhöht dies die Chance dafür, dass Fälle der falschen latenten (Wachstums-)Klasse zugeordnet werden. Untersuchungen und Vergleiche zu diesem Teilaspekt werden dadurch erschwert, dass noch kein Konsens zu der besten Definition von Klassenseparation besteht (vgl. Depaoli, 2013: 190). Simulationsstudien nutzen in der Regel die multivariate Mahalanobis Distance (MD) für latente Klassen, um eine Kenngröße der Klassenseparation anzugeben. Die Mahalanobis Distance ist hierbei durch:

$$MD = \sqrt{(\mu_1 - \mu_2)^\top \Sigma^{-1} (\mu_1 - \mu_2)} \quad (12)$$

gegeben, wobei μ_1 und μ_2 für die Mittelwertvektoren zweier latenter Klassen stehen und Σ deren gemeinsame (homogene)¹⁴ Kovarianzmatrix darstellt (vgl. McLachlan & Peel, 2000: 198). Koo & Leite (2014) zeigen anhand von Simulationen, dass eine kleine Klassendistanz (hier: $J = 1.076$; $J = 1.109$) zu Verzerrungen hinsichtlich von Parameterschätzungen, Verteilungen der Klassenzugehörigkeiten und der Identifikation der korrekten Anzahl von latenten Klassen führen (vgl. Koo & Leite, 2014: 223). Hinsichtlich der Modellkonvergenz und Parameterschätzungen können Tueller & Lubke (2010) zeigen, dass Mixture Modelle (bei gleichverteilter Klassengröße) ab einer Klassendistanz von 2.0 (MD) bereits ab 100 Fälle gute Performanz aufweisen. Bei geringerer Klassendistanz von 1.5 und 0.5 sind dagegen Fallzahlen von 300 und 1000 nötig, um zu zufriedenstellenden Ergebnissen zu gelangen (vgl. Tueller & Lubke, 2010: 189). Auch Serang et al. (2016) kommen zu dem Ergebnis, dass bei großer Klassendistanz bereits eine Fallzahl von 200 ausreichend sein kann (vgl. Serang et al., 2016: 11).

Eng verbunden mit der Klassenseparation sind die Anteilsverhältnisse in den Klassen. Hier zeigt sich, dass insbesondere Verteilungen mit einer kleinen Klasse (unter 7%) unabhängig von der untersuchten Fallzahl eine geringe Chance auf Identifikation der korrekten Klassenanzahl aufweisen (vgl. Depaoli, 2013: 190). Allerdings lassen sich diese Probleme nicht generell auf alle Abweichungen von einer Gleichverteilung der Klassengrößen erweitern.

¹⁴Wenn die latenten Klassen eine unterschiedliche Kovarianzmatrix aufweisen, muss die Klassendistanz anders berechnet werden. Die J Statistic nach James (1954) ist hierfür eine geeignete Alternative (vgl. Koo & Leite, 2014: 213).

Simulationen welche mit moderat (z. B. 75/25) ungleichgewichteten Klassen operieren, zeigen keine nennenswerten anderen Effekte auf die Modellperformanz wie die gleichgewichteten Klassenverhältnisse (vgl. Liu & Hancock, 2014: 190).

Nennenswerte Probleme bei der Schätzung von GMMs sind folglich erst bei extremen Verhältnissen bzgl. Klassenseparation und Anteilwertverhältnissen zu erwarten. Als problematisch sind allerdings auch Kombinationen aus geringer Klassenseparation, kleinen Klassengrößen und geringen Fallzahlen zu werten.

3.2.3 Konditionale Normalität der Daten

Auch wenn GMMs in der gesamten gepoolten Stichprobe keine Normalverteilung benötigen, arbeiten sie mit der Annahme von normalverteilten Daten (evtl. konditional unter Berücksichtigung von Prädiktoren) innerhalb der latenten Klassen (vgl. Bauer, 2007: 765). Abweichungen hiervon können die Identifikation der korrekten Anzahl von Klassen verzerren. Simulationen von Bauer und Curran konnten zeigen, dass nicht-normalverteilte Daten zu einer Überextraktion von latenten Klassen führen (vgl. Bauer & Curran, 2003). Hierfür generierten sie Daten, welche aus einer einzigen Population stammen, aber leicht von der Normalverteilung abweichen. Im Ergebnis zeigte sich, dass die Fit-Indizes durchgängig eine GMM Modellierung mit mehreren Klassen favorisierten, obwohl der zugrundeliegende Data Generation Process (DGP) aus einer einzelnen Population stammte. Dieses Ergebnis zeigte sich selbst für Simulationskonfigurationen, in welchen der strukturelle Part der Modelle korrekt im Sinne des DGP spezifiziert wurde (vgl. Bauer & Curran, 2004: 19). Somit können auf Grundlage der Verletzung dieser Annahme, Klassen identifiziert werden, welche keinerlei substantielle inhaltliche Bedeutung haben. Ein zuverlässiger Indikator zur Unterscheidung zwischen tatsächlich vorhandenen Subpopulationen und homogenen nicht-normalverteilten Daten aus einer Population ist weiterhin ein Forschungsdesiderat (vgl. Peugh & Fan, 2012: 210) und somit noch ein ungelöstes Problem.

3.2.4 Fallzahl

Je höher die zugrundeliegende Fallzahl bei einer GM-Modellierung, desto besser. Darin unterscheiden sich GMMs generell nicht von anderen Verfahren. Eine zu niedrig angesetzte Fallzahl kann im Rahmen von GM-Modellierungen mit Problemen hinsichtlich der Modellkonvergenz, mit unzulässigen Parameterschätzwerten wie auch mit der fehlenden Fähigkeit zur Identifikation von kleinen aber inhaltlich interessanten Subpopulationen as-

soziiert sein (vgl. Berlin et al., 2014: 14). Hinsichtlich der Klassenextraktion erhöht eine große Fallzahl die Trefferquote der korrekten Identifikation der Klassenanzahl von statistisch konsistenten Maßzahlen wie CAIC und BIC (sowie deren Derivate), wohingegen Größen wie Entropie oder AIC durch ein größeres Sample nicht unterstützt werden (vgl. Liu & Hancock, 2014: 570). Relativ einfache Simulationskonfigurationen mit wenigen latenten Klassen, adäquat hoher Klassendistanz und ausreichend großen Klassen zeigen bei einer Fallzahl ab 700 eine nahezu perfekte Klassenidentifikationsrate für Varianten von LMR und BLRT (vgl. Liu & Hancock, 2014: 570 f.). Allerdings lässt sich im Rahmen einer GM-Modellierung nur sehr problematisch eine klare Daumenregel angeben, da zahlreiche Rahmenbedingungen sowie die Interaktion miteinander eine Rolle spielen. Kim kommt anhand von Simulationsergebnissen zu dem Schluss, dass die Fallzahl umso größer sein muss,

- je höher der Missinganteil in der Stichprobe ist,
- je geringer die Trennung zwischen den latenten Klassen ausfällt,
- je mehr latente Klassen identifiziert werden sollen und
- je weniger Indikatoren (Messzeitpunkte) zur Verfügung stehen

(vgl. Kim & Kim, 2012: 474). Darüber hinaus bestehen Abhängigkeitsverhältnisse in dem Sinne, als dass auch bei einer großen Anzahl von latenten Klassen die benötigte Fallzahl nicht zwangsläufig steigen muss, wenn gleichzeitig die Anzahl der Messzeitpunkte erhöht wird (vgl. Kim & Kim, 2012: 474 f.). Da die Fallzahl im Rahmen der empirischen Anwendung noch am ehesten beeinflussbar ist, sollten bereits in der Planungsphase einer Untersuchung über plausible Werte hinsichtlich der assoziierten Größen nachgedacht werden, damit die Fallzahl adäquat angesetzt werden kann.

3.2.5 Kovariate und GMMs

Nach anfänglichen Empfehlungen GMMs möglichst mit relevanten Kovariaten zu schätzen (vgl. Li & Hser, 2011: 12 ff.; Muthén, 2004: 352 f.), zeigen neuere Studien, dass von der Berücksichtigung von Kovariaten tendenziell abzuraten ist und wenn doch, diese dann erst in einem zweiten Schritt zu berücksichtigen wären (vgl. Diallo et al., 2017: 180 ff.; Hu et al., 2017: 12; Tofghi & Enders, 2008: 335 f.). Allerdings lassen sich weiterhin Hinweise darauf ausmachen, dass bei geringer Klassendistanz und kleiner Fallzahl die Berücksichtigung

von Kovariaten von Vorteil sein kann (vgl. z. B. Hu et al., 2017: 12; Peugh & Fan, 2012: 223). Als ursächlicher Mechanismus hierfür wurde vorgebracht, dass die Einbeziehung von Kovariaten die unerklärliche Variabilität reduziert, mit welcher das Mixture-Modell arbeitet, und die Daten letztlich homogener macht (vgl. McNeish & Harring, 2017: 12).

3.2.6 Spezifikationsfehler und Invarianzannahmen

Wenn Varianz und Kovarianz der Wachstumsparameter zwischen den einzelnen latenten Gruppen als invariant angenommen werden, kann dies zu einer Überextraktion von latenten Klassen sowie zu verzerrten Parameterschätzungen führen (vgl. Bauer & Curran, 2004; Enders & Tofighi, 2008: 77; Wickrama, 2016: 204 f.). Darüber hinaus zeigen Morin et al. (2011), dass falsche Invarianzannahmen nicht nur die Klassenanzahl verzerren können, sondern auch die durchschnittliche Form der Wachstumsverläufe und die Anteilswerte der latenten Klassen (vgl. Morin et al., 2011: 638 f.). Auch wenn die Ergebnisse von Morin et al. (2011) nur auf empirischen Daten basieren und nicht systematisch im Rahmen eines Simulationsmodells untersucht wurden, liefern sie deutliche Hinweise auf potentielle Probleme von nicht geprüften Invarianzannahmen. Die Simulationsergebnisse von Diallo et al. (2016) und Enders & Tofighi (2008) zeigen zudem, dass eine freie Spezifikation von Kovarianz (und Residualvarianzen) innerhalb der Klassen generell vorzuziehen ist und so gut wie nie mit schlechterer, aber oft besserer Schätzperformanz einhergeht (vgl. Diallo et al., 2016: 20; Enders & Tofighi, 2008: 91 f.). Folglich sollte die Kovarianzmatrix für die einzelnen latenten Gruppen frei geschätzt werden. Wenn keine modell- oder modellierungsspezifischen Gründe, wie fehlende Modellkonvergenz oder klare theoretische oder analytische Gründe für Invarianz bestehen, ist dies im Rahmen der GM-Modellierung generell möglich. Allerdings erhöht sich durch diese Form der Spezifikation auch die Modellkomplexität und damit die Rechenzeit der Modelle deutlich. Darüber hinaus zeigt sich in der empirischen Praxis, dass bei einer systematischen Testung der Invarianzannahmen, selbige oft auf Grundlage der Empirie verworfen werden müssen (vgl. Diallo et al., 2016: 3).

Neben Invarianzannahmen sind noch andere Formen von fehlerhaften Spezifikationen zu berücksichtigen. Generell lässt sich dieser Aspekt der falschen Invarianzannahme verallgemeinern. Jedwede Form von Missspezifikation der Within-Klassenmatrix kann im Rahmen einer Mixture-Modellierung zu einer Überextraktion von latenten Klassen führen. Hierbei wird der fehlende Modellfit dadurch bearbeitet, dass eine zusätzliche Klasse die abwei-

chenden Fälle im Rahmen einer neuen (wenn auch evtl. künstlichen) latenten Klasse verarbeitet. Dieses Problem ist hierbei auf die elementare Interdependenz zwischen Within- und der Between-Klassenkomponente zurückzuführen (vgl. Bauer & Curran, 2004: 22). Der gleiche Mechanismus zeigt sich auch bei falschen Linearitätsannahmen in der Modellspezifikation. Bauer & Curran (2003) zeigen, dass nicht-lineare Beziehungen zwischen latenten und/oder manifesten Variablen im Rahmen eines GMMs durch die Extraktion von zusätzlichen latenten Klassen «kompensiert» werden können, auch wenn der nicht-linearen Beziehung nur eine Population zugrunde liegt (vgl. Bauer & Curran, 2004: 22 ff.). McNeish & Harring (2017) setzen zu diesen Bedenken hinsichtlich klasseninterner Fehlspezifikationen einen gewissen Kontrapunkt, indem sie zeigen können, dass Missspezifikationen insbesondere dann die Klassenidentifikation negativ beeinflussen, wenn die Anzahl der verwendeten Startwertesets zu niedrig gewählt wurde. Bei einer adäquat hohen Zahl an Startwertesets zeigen die Kennzahlen BIC und BLRT auch unter Fehlspezifikationen eine gute Performanz. Der Programmstandard von 20 Startwertesets (in Mplus) scheinen deutlich zu niedrig zu sein, wenn eine Fehlspezifikation der Wachstumsverlaufsfaktoren vorliegt, wobei die Autoren für mindestens 100 Sets plädieren.¹⁵

Eine andere Form von Fehlspezifikation werden von Chen et al. (2010) untersucht. Sie untersuchen, wie sich die fehlende Berücksichtigung von Mehrebenenstrukturen auf die Performanz von GMMs auswirkt. Ihre Ergebnisse zeigen, dass die Klassenidentifikation und die Mittelwerte der Wachstumsfaktoren kaum durch eine inadäquate Berücksichtigung der Mehrebenenstruktur beeinflusst werden. Erwartungsgemäß zeigt sich allerdings, dass die Varianzschätzungen und die Standardfehler der Wachstumsverlaufsform negativ beeinflusst werden (vgl. Chen et al., 2010: 587).

Zusammengefasst zeigen diese Ergebnisse, dass vor der eigentlichen GM-Modellierung ein geeignetes Baselinemodell gefunden werden muss, welches die wesentlichen strukturellen Eigenschaften adäquat berücksichtigt, um später schwerwiegende Fehlinterpretationen zu vermeiden (siehe Abschnitt 2.1.2). Wenn ein solches Modell gefunden wurde, sollte im Rahmen der GMM Modells zusätzlich darauf geachtet werden, dass klassenspezifische Parameter nach Möglichkeit frei zu schätzen sind und nicht a priori als invariant angenommen werden.

¹⁵Da Missspezifikationen – wenn auch nicht in den Wachstumsfaktoren – in dieser Arbeit auch Anwendung finden (bzgl. Faktorladungen und manifesten Intercepts), werden die Startwertesets auf 200 fixiert, was auch aktuellen Empfehlungen aus der Simulationsliteratur entspricht (vgl. Li et al., 2014: 510 f.).

4 Offene Fragen zu MIGMMs

In den folgenden Kapiteln 4 bis 7 werden die Forschungslücken hinsichtlich der Leistungsfähigkeit von MIGMMs herausgearbeitet und die Strategie zur Schließung dieser Lücken expliziert. In den vorangegangenen Abschnitten 2 und 3 wurden die Grundidee von Growth Mixture Modellen, deren Schätzung und Anwendungsgebiete sowie bisherige Ergebnisse zu deren Performanz herausgearbeitet. Im Weiteren gilt es auf Grundlage dieser Vorarbeit zu klären, welche offenen Forschungsfragen weiter existieren, welche Vermutungen sich aus dem bisherigen Wissensfundus ableiten lassen und mit welchem Vorgehen diese Vermutungen systematisch überprüft werden können. Zu diesem Zweck werden Arbeitshypothesen entwickelt, welche auf bisherigen Überlegungen und Performanzergebnissen zu GMMs sowie auf Ergebnissen bzgl. LGC-Modellen zweiter Ordnung basieren. Darüber hinaus wird dargestellt, wieso und auf welche Weise die Monte-Carlo-Simulationsmethodik geeignet ist, um die Performanz von MIGMMs zu untersuchen. Neben einer allgemeinen Erläuterung zur generellen Idee hinter MC-Studien wird in diesem Schritt darüber hinaus erläutert, welche Aspekte der Simulationsstudie zu diesem Zweck zu variieren sind. Zu dieser Erläuterung zählen auch die Explikation und die Rechtfertigung der Simulationsvariablen sowie deren Ausprägungen. Hierbei muss sowohl plausibilisiert werden, welche unabhängigen Einflussgrößen (wie sehr) variiert werden, welche Aspekte konstant gehalten und welche abhängigen Größen als Maß für die Performanz der Modelle herangezogen werden können.

In den diesem Abschnitt folgenden Ausarbeitungen in den Kapiteln 8 bis 11 soll neben einer detaillierten Darstellung der finalen Simulationskonfiguration und der praktischen EDV-technischen Durchführung, der eigentliche Test der in Kapitel 6 formulierten Hypothesen erfolgen. Das in Kapitel 7 aufgezeigte Analysedesign wird also in den Abschnitten 9 bis 11 durchgeführt und zugleich anhand der hier definierten Zielgrößen bewertet. Das letztendliche Ziel des Projekts ist es, Empfehlungen hinsichtlich der empirischen Forschungspraxis abzuleiten. Besonders die korrekte Identifikation von latenten Klassen im Rahmen von MIGMMs sowie deren Konditionalität bzgl. der variierten Simulationsgrößen steht hierbei im Fokus.

4.1 Desiderate bzgl. Performanz von MIGMMs

Sozialwissenschaftliche Untersuchungen arbeiten häufig mit latenten Konstrukten, welche nicht direkt durch manifeste Indikatoren messbar sind. Eine wichtige Möglichkeit dies in einer statistischen Analyse zu berücksichtigen ist ein sogenanntes Common-Factor-Modell, welches die latenten Konstrukte über reflektive Messmodelle darstellt.¹⁶ Die latenten Konstrukte werden im Prozess der statistischen Analyse wiederum zwischen verschiedenen Gruppen auf die eine oder andere Art verglichen. Eine zentrale Voraussetzung für solch eine Analyse ist, dass in unterschiedlichen Gruppen tatsächlich das Gleiche miteinander verglichen wird. Anders formuliert bedeutet dies, dass Messinvarianz hinsichtlich der latenten Konstrukte gegeben sein sollte. Die Überprüfung der sogenannten Messinvarianz zwischen Subgruppen hat im Kontext von Querschnittsanalysen eine lange Tradition und gehört schon zum Standardrepertoire für die Überprüfung von Messinstrumenten. Grundsätzlich wird dabei der Frage nachgegangen, ob einem Messinstrument für zwei (oder mehr) Gruppen die gleiche inhaltliche Bedeutung zukommt (vgl. Urban & Mayerl, 2014: 167). Falls Messinvarianz gegeben ist, kann man somit sicher sein, dass Unterschiede in den Messungen auch auf tatsächliche Unterschiede zwischen den Gruppen zurückzuführen sind. Wenn die Annahme von Messinvarianz hingegen nicht haltbar ist, lässt sich nicht mehr unterscheiden, worauf Gruppenunterschiede basieren. Es bleibt unklar, ob unterschiedliche Messwerte zwischen Gruppen auf inhaltlich substantiellen Unterschieden basieren oder lediglich ein methodisches Artefakt darstellen, welches auf die Differenzen hinsichtlich der unterschiedlichen inhaltlichen oder messtechnischen Bedeutungen zurückzuführen ist. Die Überprüfung von Messinvarianz zwischen Gruppen wurde daher auch schon extensiv untersucht (vgl. Olivera-Aguilar, 2013: 1 f.).

Die gleiche Logik des Gruppenvergleichs lässt sich auf Untersuchungen im Längsschnitt übertragen. Auch hier muss die Messinvarianz sichergestellt werden, damit gültige und zuverlässige Ergebnisse abgeleitet werden können. Wenn der Wandel hinsichtlich eines latenten Konstrukts über die Zeit untersucht werden soll, müssen Veränderungen des Messwerts auf tatsächlichen Wandel hinsichtlich der inhaltlichen Dimension hindeuten. Veränderungen der Messwerte sollten hingegen nicht eine Verschiebung der inhaltlichen oder skalentechnischen Bedeutung eines Messinstruments erkennen lassen. Die einfache

¹⁶Dabei wird für ein latentes Konstrukt ein Messmodell konstruiert, welches die Beziehung zwischen dem Konstrukt und direkt messbaren manifesten Indikatoren festlegt, wobei die manifesten Indikatoren in der Regel als messfehlerbehaftet angenommen werden (vgl. Urban & Mayerl, 2014: 128).

(ungeprüfte) Annahme, dass ein identisches Messinstrument im zeitlichen Verlauf die gleiche Bedeutung behält, reicht hierfür nicht aus. Allerdings gehen GMMs auf Basis von manifesten Indikatoren oder Composite Scores implizit von dieser längsschnittlichen Messinvarianz aus, ohne diese tatsächlich prüfen zu können. So kann, beispielsweise, im Bereich der Sozialforschung die gleiche Antwortkategorie eines Frageitems über die Zeit, unter Berücksichtigung von sozialem und individuellem Wandel von Untersuchungsobjekten, eine andere Bedeutung haben (vgl. Zhang, 2016: 5). Da die Verwendung des gleichen Messinstruments keine Garantie dafür ist, dass die Messung die gleiche Bedeutung behält, können Parameterschätzung sowie die Identifikation von Wachstumsverlaufsgruppen verzerrt werden.

Tatsächlich kombinieren sich bei GMMs diese zwei aufgezeigten Formen von Invarianzannahmen. Es wird nicht nur davon ausgegangen, dass eine längsschnittliche Invarianz hinsichtlich des Messinstrumentes besteht, sondern es wird zusätzlich unterstellt, dass sich die Messinstrumente zwischen den latenten Gruppen (welche bei GMMs identifiziert werden sollen) nicht in ihrer inhaltlichen und skalaren Bedeutung unterscheiden.

Auswirkungen und Bedeutung der Verletzung von längsschnittlichen Invarianzannahmen wurden bereits im Kontext von LGC-Modellen untersucht (z. B.: Leite, 2007; Olivera-Aguilar, 2013; Wirth, 2008). Hierbei zeigte sich, dass die Schätzung des Wachstumsverlaufs sowie die Parameterschätzer der Wachstumsfaktoren in Abhängigkeit von der Stärke der Missspezifikation entsprechend deutlicher verzerrt werden.

Auch Invarianzverletzungen im Gruppenvergleich wurden bereits im Rahmen von LGC-Modellen analysiert (vgl. Kim & Willson, 2014: 408 ff.). Kim & Willson (2014) können anhand von LGC-Multigruppen-Modellen zeigen, dass die Verletzung von (gruppenspezifischer) Invarianz die Wachstumsparameter sowie den Typ 1 Fehler verzerrt (vgl. Kim & Willson, 2014: 421 ff.).

(Second-Order/Multiple Indikator)LGC-Modelle können als ein Spezialfall von (MI)-GMMs verstanden werden. Ein (MI)GMM wird dann zu einem (MI)LGC-Modell, wenn nur eine latente Wachstumsklasse identifiziert wird, welche auch nur ein durchschnittliches Wachstumsverlaufmuster (mit zufälliger Variation) aufweist. Aufgrund dieser strukturellen Nähe der Modellklassen ist davon auszugehen, dass fehlende Messinvarianz in (MI)GMMs zu vergleichbaren Verzerrungen wie in den (MI)LGC-Modellen führen wird.

Für MIGMMs existieren, mit der Ausnahme von Zhang (2016) und einer aktuellen Veröffentlichung von Kim & Wang (2017), bisher keine systematischen Untersuchungen

der Folgen von fehlender Messinvarianz. Zhang (2016) konzentrierte sich in den Analysen auf die längsschnittliche Messinvarianz und deren Folgen für korrekte Schätzung der Wachstumsverläufe sowie auf die Genauigkeit der Klassifikation einzelner Fälle in die latenten Wachstumsklassen. Damit wurde die erste systematische Untersuchung der Folgen von fehlender Invarianz im Kontext von MIGMMs durchgeführt. Da es sich hierbei um die erste Untersuchung handelt, bleiben notwendigerweise noch viele offene Fragen übrig. So blieb der mindestens ebenso relevante Aspekt der gruppenspezifischen Messinvarianz gänzlich außen vor. Dies ist besonders beachtenswert, da die Identifikation von (latenten) Wachstumsgruppen die eigentliche Besonderheit von GM-Modellierungen darstellt. Darauf aufbauend wurde auch die Interaktion zwischen der Verletzung von gruppenspezifischer und längsschnittlicher Messinvarianz nicht berücksichtigt. Dies gilt sowohl für die Auswirkungen hinsichtlich von Parameterschätzungen aber auch bezüglich der Konsequenzen für die Identifikation der korrekten Anzahl an latenten Klassen. Allgemein wurde die Identifikation der korrekten Klassenanzahl bei Zhang (2016) nicht als Zielgröße untersucht, sondern lediglich die Präzision in der Fallzuteilung zu den latenten Klassen, was für die empirische Forschungspraxis nicht die gleiche inhaltliche Relevanz aufweist.

Kim & Wang (2017) gehen hingegen den umgekehrten Weg, indem sie sich im Kontext von MIGMMs auf fehlende Messinvarianzen zwischen latenten Gruppen konzentrieren, dabei aber den Aspekt der längsschnittlichen Non-Invarianz außen vorlassen. Sie untersuchen zwar (im Gegensatz zu Zhang (2016)) direkt die Klassenidentifikation und den Bias der Parameterschätzungen, können aber aufgrund der fehlenden Berücksichtigung von längsschnittlichen Nicht-Invarianzmustern keinen Vergleich zwischen beiden Formen anstellen und ebenso nicht die Kombination aus beiden Fällen untersuchen.

4.2 Forschungsfragen und Relevanz der Studie

In dieser Studie sollen folglich drei zentrale Forschungsfragen bearbeitet werden:

- 1) Beeinflussen gruppenspezifische und längsschnittliche Verletzungen der Messinvarianz die Fähigkeit von MIGMMs die korrekte Anzahl latenter Wachstumsverlaufklassen zu identifizieren? Wenn ja, welche Faktoren beeinflussen die Stärke dieses Effekts?

- 2) Beeinflusst die Kombination aus gruppenspezifischer und längsschnittlicher Verletzung der Messinvarianz die Fähigkeit zur Klassenidentifikation zusätzlich? Wenn ja, welche Faktoren beeinflussen die Stärke dieses Effekts?
- 3) In welchem Umfang lassen sich Parameterverzerrungen (Punktschätzer und Standardfehler) in den Wachstumsverlauffaktoren bei gruppenspezifischer, längsschnittlicher und kombinierter Verletzung der Messinvarianzannahme identifizieren? Welche Faktoren beeinflussen den Grad der Verzerrung zusätzlich?

Bisher existieren keine Untersuchungen zu kombinierten Auswirkungen von gruppenspezifischer und längsschnittlicher Verletzung der Messinvarianz im Kontext von (MI)GM-Modellen. Daher kann die Performanz von MIGMMs unter diesen Bedingungen bisher weder hinsichtlich der Klassenidentifikation noch bezüglich der Unverzerrtheit von Wachstumskurvenverläufen hinreichend bewertet werden. Darüber hinaus ist die Verletzung der Messinvarianz in der empirischen Praxis weit verbreitet und in verschiedensten inhaltlichen Kontexten präsent und potentiell problematisch. Auch ein Vergleich zwischen verschiedenen Selektionskriterien für die Klassenidentifikation (z. B. IC-Indizes vs. BLRT) fehlt bisher für MIGMMs völlig.

Diese Studie hat damit sowohl Relevanz für die Methodenforschung als auch für die empirische Forschungspraxis, welche sich vorrangig mit inhaltlichen Fragen befasst. Die Ergebnisse dieser Studie sollen der empirischen Praxis Orientierung liefern, indem der Grad des Vertrauens, welchen Forscher/Forscherinnen in die Ergebnisse von MIGMMs legen können, systematisch evaluiert wird. Die Relevanz der Überprüfung von alternativen Modellierungen, welche, beispielsweise, fehlende Messinvarianzen berücksichtigen, wird dabei herausgearbeitet.¹⁷ Im Idealfall lassen sich anhand der Ergebnisse eindeutige Richtlinien formulieren, welche bei der Modellierung von MIGMMs berücksichtigt werden sollten.

Um die hier dargestellten Wissenslücken zu bearbeiten und den damit assoziierten Forschungsfragen nachzugehen, wird in dieser Studie eine Monte-Carlo-Simulation durchgeführt.

¹⁷Hierunter fällt, beispielsweise, die Post-Hoc-Validierung von Messmodellen und Wachstumsverläufen der identifizierten latenten Klassen. So ließe sich herausarbeiten, unter welchen Bedingungen derlei Zusatzanalysen besonders von Nöten werden, da mit potentiell starken Verzerrungen zu rechnen ist.

5 Prinzipien für Monte Carlo Studien

Monte-Carlo-Studien weichen in ihrer Zielsetzung deutlich von empirisch statistischen Untersuchungen ab und gehören noch nicht zum standardisierten Methodenrepertoire der Sozialforschung. In ihrer Anwendung als Untersuchungsinstrument von Analyseverfahren gilt es einige idiosynkratische Aspekte hervorzuheben, um die Nachvollziehbarkeit für den interessierten Leser zu gewährleisten. Daher wird im Folgenden zunächst die Grundidee von MC-Studien erläutert, um daran anschließend Besonderheiten bei der Dokumentation, Durchführung und Ergebnisinterpretation zu verdeutlichen.

5.1 Grundidee von Monte Carlo Studien

Allgemein kann immer dann von einer Monte-Carlo-Simulation gesprochen werden, wenn statistische Experimente wiederholt (bzw.: repliziert) und die Ergebnisse dieser Experimente gezielt analysiert werden. Computergenerierte Zufallszahlen werden dabei genutzt, um verschiedene Szenarien eines Experiments zu simulieren (vgl. Falk et al., 2014: 296). Eine Monte-Carlo-Simulation beinhaltet, typischerweise, die Ziehung von einer Vielzahl von Zufallsstichproben (Replikationen¹⁸). Diese Ziehung geschieht computergesteuert aus einer vorher klar festgelegten Population, welche durch einen sogenannten Data Generating Process (DGP) mathematisch genau in seinen systematischen und stochastischen Komponenten definiert ist. Diese Zufallsstichproben werden in der Gesamtbetrachtung nach Muster untersucht, wobei diese Muster wiederum mit der wahren Population (sprich der vorher festgelegten Wirklichkeit) verglichen werden können (vgl. Carsey & Harden, 2014: 7). In weiteren Schritten kann der DGP modifiziert werden, um zu untersuchen, ob diese Modifikation zu anderen Mustern in den Stichproben bzw. zu anderen Vergleichssituationen zwischen DGP und den Mustern der Stichproben führt. Im Unterschied zur klassischen empirischen statistischen Untersuchung in den Sozialwissenschaften ist bei MC-Studien der wahre DGP bekannt, da er vom Forscher festzulegen ist. Der DGP ist nicht das angestrebte Endziel der statistischen Analyse, sondern beschreibt im vornhinein und in mathematisch darlegbarer Form, welche Zusammenhänge und Muster in den Referenz-Daten bestehen. Aus dieser so beschriebenen Datenrealität werden im nächsten Schritt Stichproben gezogen. Die Ergebnismuster dieser Stichproben werden in der Gesamtbe-

¹⁸Die Anzahl an Replikationen kann im Rahmen einer MC-Studie konzeptionell als Fallzahl verstanden werden (vgl. Kim & Kim, 2012: 462), wobei die eigentliche Fallzahl (bzw. Sample Size) oft eine simulierte Größe darstellt.

trachtung gegenüber dem DGP bewertet, welcher hier als Referenz dient. Anders als bei der Beobachtung oder dem Experiment ist der Zugang zur Realität, den tatsächlichen Mustern in der Population, bei der Simulationsstudie also bereits vorhanden. Ziel ist folglich nicht der Inferenzschluss von identifizierten Maßen und Zusammenhängen in der Stichprobe auf eine irgendwie geartete Grundgesamtheit. Bei der Monte-Carlo-Simulation steht in der Regel im Vordergrund, wie sich Veränderungen/Gegebenheiten in der Population oder bei der Stichprobenkonstruktion auf unsere Schlüsse über die Population auswirken. Somit ist das Ziel von MC-Studien im Vergleich zu empirischen Studien nicht der Inferenzschluss auf die Grundgesamtheit, sondern der Schluss auf die (oder die Bewertung von der) Qualität der Inferenzschlüsse an und für sich. Bereits Mooney (1997) stellte in einem exemplarischen Überblick zu Anwendungsformen von MC-Studien für die Sozialwissenschaften heraus, dass vor allem die statistische Inferenz (bei: fehlender statistischer Theorie¹⁹; Robustheit unter Annahmenverletzungen; bzgl. Umfang von Typ 1 (u. Typ 2) Fehlern) und die Bewertung von Koeffizientenschätzern (bei: Annahmenverletzung oder im Vergleich zwischen verschiedenen Methoden) zentrale Anwendungsfälle für den Einsatz von Monte Carlo Verfahren darstellen (vgl. Mooney, 1997: 66 ff.). Allerdings müssen Monte-Carlo-Simulationen nach der hier aufgestellten Definition von sogenannten Resampling-Methoden wie jackknifing, bootstrapping, cross validation oder permutation tests unterschieden werden. Diese Verfahren teilen zwar das Grundprinzip des Monte-Carlo Verfahrens von der wiederholten iterativen Ziehung von Stichproben, aber eben anhand einer empirischen Stichprobe/Datenlage und nicht auf Grundlage eines vorher vom Forscher klar definierten DGP, welcher als objektiver und unverzerrter Vergleichsmaßstab dienen kann (vgl. Carsey & Harden, 2014: 7).

MC-Simulationen können als eine Form von numerischen Experimenten im (Daten-)Labor verstanden werden. Alle möglichen Aspekte, welche das Ergebnis beeinflussen können, sind kontrolliert. Dadurch lässt die Manipulation einzelner Größen den kausalen Schluss auf das Ergebnis dieser Manipulation viel klarer zu, da andere Erklärungsansätze aufgrund der kontrollierten und invarianten Randbedingungen ausgeschlossen werden können. Allerdings haben MC-Studien auch die gleichen Probleme wie Laborexperimente. Die externe Validität der Ergebnisse ist nicht sicher. Diese Idee der fehlenden externen Validität lässt sich unmittelbar auf das MC-Setting übertragen. So bleibt es oft unklar, ob

¹⁹Im Bereich der Strukturgleichungsmodellierung finden (beispielsweise) Fit-Indizes wie der Comparative Fit Index (CFI) Anwendung, welcher auf keiner bekannten statistischen Verteilung beruht und dessen Grenzwerte anhand von Simulationsstudien gewonnen wurden (vgl. Bandalos & Gagne, 2012: 93).

die simulierten Bedingungen des Data Generating Processes (DGP) tatsächlich auf reale empirische Datensituationen zu übertragen sind oder ob die Muster in den Daten zu sehr stilisiert und/oder vereinfacht wurden, sodass die gezogenen Schlüsse zwar im „Datenlabor“ korrekt sind, aber in der Realität keine Relevanz haben, da die Laborsituation in der empirischen Praxis nie vorzufinden ist. Auch eine Extrapolation der im Experiment identifizierten Ergebnisse bleibt unsicher, wenn die Bedingungen des Extrapolationsbereichs in der Simulation nicht berücksichtigt wurden oder außerhalb des Bereichs der simulierten Kategorienschemas fallen. MC-Studien müssen dann Anwendung finden, wenn keine analytische Lösung existiert oder die statistischen Eigenschaften von Schätzern (in bestimmten Datensituationen) unbekannt sind. Darüber hinaus können MC-Studien auch eingesetzt werden, wenn eine analytische Lösung zwar prinzipiell möglich, aber für ein spezifisches Problem zu kompliziert oder aufwendig in der Umsetzung ist.

5.2 Durchführung und Darstellung von MC-Studien

Da MC-Studien hinsichtlich ihrer Konzeption von empirisch statistischen Untersuchungen abweichen, gilt es auch andere Kriterien für die Untersuchungskonzeption, Darstellung und Ergebnisinterpretation anzulegen. Boomsma et al. (2012) liefern hierzu einen umfangreichen Anforderungskatalog, welcher auch in den folgenden Ausführungen zur Konfiguration der Simulation sowie der Ergebnispräsentation berücksichtigt werden soll. In Anlehnung an Boomsma et al. (2012) gilt es dabei folgende Aspekte zu berücksichtigen:

1. Die Wahl der Population bzw. Populationsparameter (die simulierte Wirklichkeit) sollte klar definiert und dargestellt werden. Die Auswahl des Populationsmodells beeinflusst unmittelbar die externe Validität der Ergebnisse der Untersuchung. Darüber hinaus geben die Populationsparameter den Rezipienten der Untersuchung Anhaltspunkte bzgl. der Relevanz der Ergebnisse für eigene Studien. Wenn als Grundlage der MC-Studie ein Analysemodell gewählt wird, welches von der Forschungspraxis zu weit entfernt ist, dann ist die Übertragbarkeit der Ergebnisse generell schwieriger als bei einem Modell, welches in der empirischen Praxis eher die Regel darstellt. Daher sollten sich die Werte des zugrunde liegenden Populationsmodells an realen Szenarien aus der substanziellen empirischen Forschung orientieren.
2. Vor der Durchführung der Studie gilt es geeignete Kriterien für Performanz von Modellschätzungen und Cut-Off-Points für gute Performance festzulegen. Dies ent-

spricht der Bestimmung der abhängigen Y-Variable(n) der Simulation. Je nach Forschungsfrage gilt es individuell zu bestimmen, welche Unterschiede zwischen dem DGP und der Zusammenfassung der Schätzergebnisse einer Simulationskonfiguration betrachtet werden und ab welchem Schwellenwert die Unterschiede als akzeptabel bzw. nicht mehr akzeptabel zu bewerten sind. Beispielsweise könnte der Parameter-Bias betrachtet werden, welcher ab einer relativen Abweichung gegenüber dem Parameter des DGP von über 10 % als nicht mehr akzeptabel eingestuft wird.

3. Neben den endogenen Größen der Simulation gilt es auch die exogenen Variablen der Simulation zu bestimmen. Diese Variablen werden im Rahmen von MC-Studien auch als Design Conditions, experimentelle Faktoren oder als Variationen der Modelle bezeichnet. Sie stellen die X-Variable der Simulation dar. Im ersten Schritt geht es hierbei darum, die klare Darstellung der variierten Faktoren und deren Beziehung zueinander (bzw. deren Interaktion miteinander) aufzuzeigen, welche sich aus der Forschungsfrage der MC-Simulation ergeben. Im zweiten Schritt gilt es auch die jeweiligen Ausprägungskategorien bzw. Levels der unabhängigen Größen zu begründen. Wird die Fallzahl als unabhängige Größe einer Simulation untersucht, gilt es also dies darzustellen, Kombinationen der Fallzahl mit anderen unabhängigen Größen zu benennen und zu begründen, wieso (z. B.) die Ausprägungen 200, 500 und 1000 gewählt wurden. Die wichtigsten Ankerpunkte für die Begründung der Levels von unabhängigen Variablen sind die externe Validität, die Anschlussfähigkeit an vorangegangene Simulationsstudien sowie forschungsökonomische Beschränkungen von Simulationsstudien.²⁰
4. Sofern irgendwie möglich sollten Erwartungen hinsichtlich der Richtung und Stärke von Wirkbeziehungen zwischen unabhängigen und abhängigen Variablen noch vor der tatsächlichen Simulation formuliert werden. Ansatzpunkte hierfür können sowohl theoretische und/oder analytische Vorüberlegungen sein als auch Ergebnisse aus vorangegangenen Simulationsstudien. Bezüglich der Beziehung zwischen unabhängigen und abhängigen Variablen sollte vor allem die Effektstärke berichtet werden. Das Signifikanzniveau ist bei Simulationsstudien häufig nicht sehr aussagekräftig, da die

²⁰So ist es zwar generell wünschenswert so viele Kategorien einer unabhängigen Variable wie möglich zu simulieren, aber jede weitere Kategorie bedeutet in der Logik der Simulation weitere Simulationskonfigurationen (insbesondere bei faktoriellen Analysedesigns, bei welchem jede Kategorienausprägung mit jeder Anderen zu kreuzen ist). Dies bedeutet, dass sich die Laufzeit der Simulation hierdurch exponentiell erhöhen kann.

Anzahl der Replikationen vom Forscher prinzipiell beliebig manipulierbar ist, wodurch die statistische Power nahezu nach Wunsch gesteigert werden kann und jedes Ergebnis unabhängig von seiner substanziellen Bedeutung signifikant wird.

5. Da MC-Studien neben den klar definierten Populationsparameter auch in aller Regel Zufallskomponenten enthalten, gilt es neben dem DGP auch Startwerte und Seeds für die Replikation der Studie zu berichten. Ein nicht zu unterschätzender Aspekt ist die Dokumentation und Rechtfertigung der Wahl, der Versionsnummer und der Syntax des jeweiligen Statistikpakets (z. B. in Appendix oder Web). Dies ermöglicht nicht nur die Replikation der Ergebnisse, sondern erhöht auch die Anschlussfähigkeit für weitere Analysen, welche zusätzliche oder erweiterte Aspekte untersuchen wollen.
6. Da bei der Simulation von verschiedenen Datensituationen eine sehr große Anzahl von Ergebnissen generiert wird ²¹, müssen zur Analyse und Darstellung der Simulationsergebnisse besondere Schritte unternommen werden. So sollten die Ergebnisse immer bzgl. der Hypothesen aufgearbeitet werden, um die Maße der Daten zu ordnen. Die Vielzahl an Informationen, Maßen und Indizes gilt es dabei so zu reduzieren, dass die wesentlichen Folgerungen herausdestilliert werden können. Insbesondere tabellarische Zusammenfassungen und grafische Darstellungen sind in diesem Zusammenhang hilfreich, solange sie klar auf die Forschungsfragen bezogen werden (vgl. auch Boomsma et al., 2012: 353).

5.3 Konzept der Umsetzung der MC-Studie

Für die technische Umsetzung der Simulationsstudie werden mehrere Statistiksoftwarepakete verwendet. Die Modellschätzung wird anhand der simulierten Stichproben in Mplus (Version 7.11 Linux) durchgeführt. Auch die Generierung des DGP wird in Mplus umgesetzt. Für die Erzeugung der hierfür nötigen Inputfiles wird auf die aktuelle Version von R (Version 3.11) und das damit verbundene Softwarepaket MplusAutomation (Version 0.6-4) zurückgegriffen. Auch die Auswertung im Sinne von deskriptiv- und inferenzstatistischen Analysen der einzelnen Simulationskonfigurationen und deren Ergebnisse wird in R durchgeführt.

²¹Die Menge der generierten Ergebnisse orientiert sich dabei an dem Produkt aus der Anzahl an Replikationen, der Anzahl verschiedener Simulationskonfigurationen sowie der Anzahl der individuellen Parameter einer Modellschätzung (plus zusätzliche aggregierte Größen).

Mplus wird als Kernstück der Statistiksoftware verwendet, da es sich im Bereich der Strukturgleichungsmodellierung und insbesondere für GMMs als Quasistandard durchgesetzt hat. Somit wird in dieser Studie auch auf Mplus gesetzt, um die Anschlussfähigkeit an die empirische Praxis zu gewährleisten. Die Berechnungen selbst werden auf dem Parallelrechner BwUniCluster²² des Projekts Hochleistungsrechnen in Baden-Württemberg (bwHPC) durchgeführt, um die Laufzeit der Simulation möglichst gering zu halten. Hierbei wird jede einzelne Simulationskonfiguration als eigenständiger Rechenprozess an den Parallelrechner übergeben. Die hierfür nötigen Befehle und Parameterübergaben werden durch ein eigenes Shellskript für jede einzelne Konfiguration erstellt. Zusätzlich wird dabei aber nicht nur nach Simulationskonfiguration geteilt, sondern auch nach der jeweiligen unterstellten Anzahl an latenten Klassen. Da für eine GM-Modellierung nicht nur ein Modell nötig ist, sondern für jede mögliche Anzahl latenter Klassen ein eigenes Modell berechnet werden muss, ist dies erforderlich. Somit ergibt sich die Anzahl der Rechenmodelle aus dem Produkt der Anzahl von Simulationskonfigurationen und der Anzahl an überprüften latenten Klassen.²³ Da keine Abhängigkeit bzgl. der einzelnen Simulationskonfigurationen besteht, ist es ohne weiteres möglich die Berechnungen zu trennen. Wenn man von der Wartezeit für Rechenprozesse absieht, reduziert sich die Komplexität und Dauer der Simulation dadurch auf den Umfang einer einzelnen Konfiguration mit einer spezifischen unterstellten Anzahl an latenten Klassen.

6 Thesen/Annahmen bzgl. variierter Modellparameter

6.1 Exkurs: Wachstumsmodelle und Messinvarianzen

Es können mehrere Typen von Messinvarianz unterschieden werden, welche aufeinander aufbauen und einen unterschiedlichen Grad an Restriktionen für die Messmodelle im Längsschnittkontext verlangen. Die folgenden Ausführungen beziehen sich lediglich auf die in dieser Studie untersuchten Formen, welche für die Konstruktion von MI/SO-LGC-Modellen und in deren Erweiterung in MI/SO-GMMs relevant sind.²⁴

²²Das BwUniCluster verfügt zum aktuellen Zeitpunkt über 520 nutzbare Rechenkonten mit einer Prozessortaktung von 2,4 bis 2,6 GHz (vgl. O.V., 2017).

²³Allerdings wird hierbei unterschlagen, dass jede Modellschätzung mit mehreren Startwerten durchgeführt wird, um lokale Maxima zu vermeiden. Die tatsächliche Anzahl an geschätzten Modellen muss daher prinzipiell noch mit der Anzahl an Startwerten multipliziert werden. Für unterschiedliche Startwerte wird jedoch kein eigener Rechenprozess übergeben.

²⁴Für umfangreichere Ausführungen zu verschiedenen Messinvarianzformen siehe Urban & Mayerl, 2014: 167 ff. oder Wirth, 2008: 19 ff.

Wenn Wachstumsverläufe von latenten Konstrukten untersucht werden sollen, dann ist diese Untersuchung nur dann sinnvoll, wenn die Faktorladungen und manifesten Itemintercepts im Längsschnitt als invariant angenommen werden (vgl. Brown, 2006: 257). Chan (1998) unterscheidet in diesem Kontext zwischen Alpha-, Beta- und Gamma-Wandel, was die Messproblematik bei der Analyse von Wachstumsverläufen sinnvoll strukturiert. Nur der Alpha-Wandel beschreibt eine tatsächliche zeitliche Veränderung hinsichtlich des wahren Wertes eines (latenten) Konstrukts, welche allerdings nur bei Messinvarianz hinsichtlich der Messskala und der konzeptuellen Bedeutung des latenten Konstrukts nachgewiesen werden kann (vgl. Chan, 1998: 426). Ein Beta-Wandel beschreibt beobachtete Veränderungen, welche auf eine zeitliche Rekalibrierung (z. B. Streckung oder Schrumpfung) der Messskala zurückzuführen sind, wohingegen der Gamma-Wandel auf einer zeitlichen Verschiebung der Bedeutung des untersuchten Konstrukts beruht (vgl. Chan, 1998: 426; Brown, 2006: 253). Ein tatsächlicher Alpha-Wandel kann folglich nur dann untersucht werden, wenn kein Beta- oder Gamma-Wandel vorherrschen. Darauf aufbauend müssen verschiedene Formen von Messinvarianz beachtet werden, welche sicherstellen, dass nur ein tatsächlicher Wandel inhaltlich interpretiert wird.

Konfigurale Invarianz. Die basalste Form der längsschnittlichen Messinvarianz wird als konfigurale Faktorinvarianz bezeichnet. Diese Invarianzform beinhaltet, dass jede Messung die gleichen Items enthält und die relativen Faktorgewichte sich über die Zeit betrachtet nicht verändern, wodurch die Beziehungsstruktur der Faktorladungen zu jedem Erhebungszeitpunkt erhalten bleibt (vgl. Urban & Mayerl, 2014: 168 f.). Die konstante Rangordnung der Faktorgewichte betrifft nach der Unterscheidung nach Chan (1998) den Gamma-Wandel, welcher hierdurch eingeschränkt wird.

Metrische Invarianz. Eine darauf aufbauende und fortgeschrittenere Form der Messinvarianz ist die metrische Faktorinvarianz.²⁵ Hierfür wird verlangt, dass die unstandardisierten Faktorladungen eines jeden Messindikators über die Zeit/Gruppe konstant bleiben (vgl. z. B. Urban & Mayerl, 2014: 169). So wird für jede (unstandardisierte) Faktorladung (λ) jedes manifesten Items i verlangt, dass für jeden Messzeitpunkt t gilt:

$$\lambda_{i,t=1} = \dots = \lambda_{i,t+n} \quad (13)$$

²⁵Alternative Bezeichnungen: weak invariance; factor pattern invariance (vgl. z. B. Wirth, 2008: 20).

Diese Form der Messinvarianz stellt sicher, dass die latenten Konstrukte im Wandel der Zeit die gleiche konzeptuelle inhaltliche Bedeutung beibehalten. Auf der Ebene des Gruppenvergleichs bedeuten gleiche Faktorladungen, dass ein Wandel um eine Messeinheit in einer Gruppe äquivalent zu einem Wandel um eine Messeinheit in einer anderen Gruppe ist und somit diesbezüglich vergleichbar bleibt (vgl. Wang et al., 2018: 170). Damit betrifft die metrische Invarianz in der Aufschlüsselung nach Chan (1998) ebenfalls die Vermeidung von Gamma-Wandel.

Skalare Invarianz. Die sogenannte starke oder skalare Invarianz geht hinsichtlich der Restriktionen der Messmodelle noch einen Schritt weiter. Bei dieser Form der Messinvarianz wird zusätzlich gefordert, dass die Intercepts eines jeden manifesten Indikators zeitlich/gruppenspezifisch konstant bleiben (vgl. z. B. Wirth, 2008: 22). So wird verlangt, dass nicht nur die Annahme der metrischen Faktorinvarianz gilt, sondern zusätzlich für jeden Item-Intercept (τ) jedes Item i zu jedem Erhebungszeitpunkt t jeweils die Gleichungen:

$$\begin{aligned}\tau_{i,t=1} &= \dots = \tau_{i,t=t+n} \\ \lambda_{i,t=1} &= \dots = \lambda_{i,t=t+n}\end{aligned}\tag{14}$$

erfüllt sind. Damit wird sichergestellt, dass ein beobachteter Wandel des latenten Konstrukts auch wirklich auf eine Veränderung des Konstrukts und nicht auf einen Wandel (oder Unterschiedlichkeit im Gruppenfall) der Messung selbst zurückzuführen ist. Hier hinter steht der Gedanke, dass jeder Unterschied bzgl. der Mittelwerte der manifesten Variablen (zwischen Gruppen oder über die Zeit hinweg) auf einen tatsächlichen Unterschied bzgl. des latenten Konstrukts zurückzuführen ist (vgl. Millsap, 2011: 105). In der Aufschlüsselung nach Chan (1998) entspricht dies am ehesten dem Beta-Fehler, welcher auf diesem Wege ausgeschlossen werden soll.

Die skalare und metrische Invarianz werden im Rahmen dieser Studie manipuliert, um den Einfluss von nicht kontrollierten Verletzungen auf der Längsschnitt- und Gruppen-Ebene auf die Performanz von MIGMMs zu untersuchen.

Absolute Messinvarianz von latenten Konstrukten in den Sozialwissenschaften mit Unterschieden in der Itemfunktion von 0.000 ist aus einer empirischen Perspektive wohl kaum möglich. Nicht die dichotome Einteilung in Messinvarianz und in Messungleichheit ist daher praktisch relevant, sondern das Ausmaß der Abweichung vom Ideal und wie

sich diese Abweichung hinsichtlich der inhaltlichen Forschungsfrage auswirkt. Daher sollte man die vollständige Messinvarianz auch nicht zu vorschnell als harte Vorbedingung für einen Gruppenvergleich bzgl. latenter Größen/Klassen anführen (vgl. auch Borsboom, 2006: 178). Eine praktikable und notwendige Vorbedingung für valide und zuverlässige Gruppenvergleiche oder Gruppenidentifikationsverfahren (wie MIGMMs) sollte allerdings die Untersuchung des Grades der Messinvarianz sowie die Analyse von Auswirkungen von Invarianzverletzungen sein. Hier setzt auch diese Studie an.

6.2 Thesen/Erwartungen bzgl. Messinvarianzen in MIGMMs

Aufbauend auf den Ergebnissen für LGC-Modelle ist zu erwarten, dass eine Verletzung der längsschnittlichen Messinvarianz sowohl die Schätzung der Varianzen als auch die Parameterschätzer der Wachstumsparameter verzerren wird (vgl. z. B.: Wirth, 2008: 106 ff.). Während Leite (2007) diese Abweichungen lediglich als moderat einschätzt (vgl. Leite, 2007: 606), identifiziert Wirth (2008) wesentlich stärkere Verzerrungen in den Varianzen und Punktschätzungen der Wachstumsparameter (vgl. Wirth, 2008: 106 ff.).

Diese unterschiedlichen Einschätzungen lassen sich wohl größtenteils auf die unterschiedliche Form der Modellierung der jeweiligen Invarianzverletzungen zurückführen. Während Leite (2007) die Invarianzverletzungen zufällig variierte, ohne zeitliche Dynamik, arbeitete Wirth (2008) mit Konfigurationen, welche einen zeitlichen Trend enthalten. Auch die Ergebnisse von Zhang (2016) (bzgl. MIGMMs) lassen darauf schließen, dass die Stärke der Invarianzverletzung das Ausmaß der Verzerrungen maßgeblich beeinflusst (vgl. Zhang, 2016: 101 f.). Damit kann Zhang (2016) die Ergebnisse von Wirth (2008) und Olivera-Aguilar (2013) (bzgl. LGC-Modellen) dahingehend bestätigen. Auch Kim & Kim (2012) zeigt anhand von LGC-Mehrgruppen-Modellen, dass die Verletzung von (gruppenspezifischer) Invarianz die Wachstumsparameter sowie den Typ 1 Fehler verzerrt. Dabei konnte gezeigt werden, dass die Art der fehlenden Invarianz entscheidend ist. Wenn die fehlende Messinvarianz ignoriert wurde, zeigten sich für Faktorladungen Verzerrungen in den Slope-Wachstumsparametern. Bei fehlender Invarianz bezüglich der Intercepts traten hingegen Verzerrungen in den Interceptfaktoren auf. Dieses Ergebnis kann durch Kim & Wang (2017) auch im Kontext von fehlender Beachtung von Non-Invarianz bei MIGMMs repliziert werden (vgl. Kim & Wang, 2017: 15 f.).²⁶ Auch eine Typ-1-Fehler-Inflation

²⁶Bei diesen Ergebnissen kann problematisiert werden, dass die bewusst vereinfachten Simulationsbedingungen (vgl. Kim & Wang, 2017: 16) (insbesondere die fehlenden diachronen Fehlerkorrelationen) von

konnte aufgrund fehlender Beachtung der Nicht-Invarianz identifiziert werden. Je stärker die fehlende Invarianz ausgeprägt war, desto stärker waren auch die Verzerrungen der Wachstumsparameter und des Typ 1 Fehlers (vgl. Kim & Willson, 2014: 421 ff.).

Zwar wurde das Zusammenspiel zwischen längsschnittlicher und gruppenspezifischer Invarianz bisher noch nicht untersucht, aber intuitiv lassen sich zwei mögliche Szenarien der Wirkung ausmachen. Bei einer gleichförmigen Effektrichtung der Verzerrung muss davon ausgegangen werden, dass sich die kombinierten Invarianzverletzungen gegenseitig aufaddieren oder möglicher Weise sogar verstärken. Falls die Invarianzverletzungen allerdings in die entgegengesetzte Richtung weisen, kann bei einer Kombination evtl. sogar von einer Kompensation der verzerrenden Wirkung ausgegangen werden, wodurch das Schätzergebnis evtl. näher am DGP dran ist als in Simulationskonfigurationen, welche lediglich eine Form der Invarianzverletzung aufweisen. Die spekulative Natur dieser Annahmen und ihre entgegengesetzte Stoßrichtung unterstreichen nur die Notwendigkeit, diese Szenarien systematisch in Rahmen einer Simulationsstudie zu untersuchen.

Für die Fähigkeit zur korrekten Klassenidentifikation bei nicht beachteter längsschnittlichen Non-Invarianz der Messung können nur wenige Ableitungen aus bisherigen Forschungsergebnissen gemacht werden, da dies im Kontext von MIGMMs bisher noch nicht explizit untersucht wurde. Zhang (2016) untersuchte für MIGMMs lediglich den Einfluss von längsschnittlichen Invarianzverletzungen auf die Klassifikationsrate auf Anteils- und Individualniveau. Hierbei zeigte sich, dass die Identifikation der korrekten Anteilswerte der latenten Klassen sowie die individuelle Fallzuordnung in die latenten Klassen nicht durch Verletzungen der längsschnittlichen Messinvarianz beeinflusst wurden (vgl. Zhang, 2016: 105 ff.). Allerdings ist dieses Ergebnis mit Vorsicht zu interpretieren, da alle verwendeten Simulationskonfigurationen insgesamt von einer großen Klassendistanz ausgegangen sind und somit kritische Schwellenwerte evtl. nicht überschreiten konnten.²⁷ Außerdem konnten keine evtl. verstärkenden Effekte aus der Kombination aus längsschnittlichen und gruppenspezifischen Invarianzverletzungen berücksichtigt werden, da dies nicht Teil der Simulationskonfiguration war. Nun sind Klassenidentifikation und Klassifikationsraten nicht das Gleiche und die Ergebnisse können nicht direkt ineinander überführt werden. Allerdings liegt die Vermutung nahe, dass die Verletzung von Invarianzannahmen insbesondere dann für die Klassenidentifikation relevant wird, wenn die latenten Klassen nahe

realen Datensituationen i. d. R. abweichen.

²⁷Selbst Zhang (2016) empfiehlt, dass die Ergebnisse unter einer stärkeren Variation der Klassendistanz repliziert werden sollten (vgl. Zhang, 2016: 110).

zusammen sind, wodurch zwei tatsächlich distinkte latente Wachstumsklassen lediglich als größere Varianz innerhalb einer zusammengefassten Klasse interpretiert werden können. Für gruppenspezifische Messunterschiede kann nach den Ergebnissen von Kim & Wang (2017) erwartet werden, dass die Klassenidentifikation der IC-Maße durch nicht beachtete Messunterschiede beeinflusst wird (vgl. Kim & Wang, 2017: 9). Allerdings darf bei dieser Beeinflussung nicht nur an die negative Richtung von weniger korrekten Identifikationsraten gedacht werden. Kim & Wang (2017) können in ihrer Simulationsstudie für eine Teilmenge der Konfigurationen erkennen, dass nicht beachtete Messunterschiede die Identifikationsraten (zumindest des BICs) begünstigen können, indem zwischen den latenten Gruppen mehr Heterogenität eingeführt wird, was wiederum die Klassenidentifikation begünstigt (vgl. Kim & Wang, 2017: 9 ff.).

6.3 Thesen/Annahmen bzgl. weiterer Modellgrößen

Neben Verletzungen von Invarianzannahmen werden auch die Modellgrößen Fallzahl, Klassenseparation und Mixing Proportions untersucht. Auch wenn das primäre Forschungsinteresse nicht auf dem Einfluss dieser Größen liegt, werden sie im Rahmen dieser Studie als Kontrollvariablen und als mögliche Quellen von Moderatoreffekten untersucht. Für die Haupteffekte dieser Modellgrößen werden ähnliche Ergebnisse erwartet, wie in vorangegangenen Simulationsstudien bzgl. GMMs (siehe hierzu ausführlicher Abschnitt 3.2). So wird angenommen, dass eine geringe Fallzahl, latente Klassen mit geringen prozentualen Anteilswerten und eine geringe Klassendistanz die Performanz von MIGMMs negativ beeinflussen werden. Diese Thesen lassen sich nach dem aktuellen Forschungsstand auf alle drei hier untersuchten abhängigen Performanzgrößen (vgl. Abschnitt 7.1) übertragen. Ggf. können die im Rahmen von Simulationsstudien mit GMMs untersuchten Effekte von Fallzahl, Klassenseparation und Mixing Proportion durch diese Untersuchung auf MIGMMs verallgemeinert werden.

Hinsichtlich der Interaktion mit Messinvarianz-Simulationsgröße wird erwartet, dass geringe Fallzahlen, kleine Klassen und geringe Distanzen zwischen den latenten Klassen den Einfluss der negativen Effekte von Verletzungen der Messinvarianz auf die Modellperformanz verstärken werden. Die schlechtesten Ergebnisse sind für Simulationskonfigurationen zu erwarten, in welchen sich die ungünstigen Bedingungen aufaddieren. Für die Klassendistanz wird darüber hinaus angenommen, dass sie die deutlichste Moderatorwirkung auf den Einfluss von Nicht-Invarianzen auf die Klassenidentifikation aufweist,

da Klassenüberschneidungen durch Messungs- bzw. Modellierungsfehler unter geringer Klassendistanz generell wahrscheinlicher werden oder (wie bei Kim & Wang (2017) gesehen) fehlende Klassendistanz durch zusätzliche Heterogenität von fehlender Messinvarianz kompensiert werden kann.

7 Konfiguration der Simulation

7.1 Zielgrößen/Abhängige Variablen

Die abhängigen Variablen der Studie stellen die Zielgrößen der Simulation dar. Hierbei gilt es zu betrachten, welche Auswirkung die unabhängigen Variablen auf das Verhältnis zwischen Kennwerten der simulierten Stichproben im Vergleich zu den korrekten Populationswerten haben. Der Grad und die Form der Abweichung zwischen Werten der Stichproben und der Population werden als Kriterien herangezogen, anhand welcher die Güte des Schätzverfahrens abgeleitet wird. Je höher die Übereinstimmung zwischen den Kennzahlen der Population und der simulierten Stichproben ausfällt, desto besser ist die Performanz des Analyseverfahrens zu bewerten.

Extraktion der korrekten Anzahl von latenten Klassen. Da die Klassenextraktion das Hauptanwendungsgebiet von GMMs darstellt, ist die korrekte Identifikation tatsächlich vorhandener Klassen das wichtigste und am häufigsten untersuchte Zielkriterium für GMMs. Die Identifikation der latenten Klassenanzahl einer simulierten Stichprobe wird hierbei anhand mehrerer Kriterien durchgeführt. Für diese Arbeit wird auf die gängigen IC-Maße AIC, BIC und aBIC zurückgegriffen. Aus dem Bereich der modifizierten Likelihood Ratio Tests wird der BLRT verwendet. Die hierüber identifizierte Klassenanzahl der simulierten Stichproben wird dann der korrekten Anzahl an latenten Klassen aus dem DGP gegenübergestellt. Dadurch können Unter- oder Überschätzungen der Anzahl an tatsächlich vorhandenen latenten Klassen identifiziert werden.

Da sich die hier verwendeten Kennzahlen der Klassenidentifikation widersprechen können, gilt es jeweils eine separate Auswertung vorzunehmen. Damit kann zugleich eine Performanzbewertung der gängigen Identifikationsgrößen für die Klassenanzahl im Kontext von MIGMMs vorgenommen werden.

Bias in den Parameterschätzungen. Neben der Extraktion der korrekten Klassenanzahl sind, des Weiteren, Verzerrungen hinsichtlich der Wachstumsparameter (Varianz und Mittelwerte) von substanziell inhaltlichem Interesse. Gleiches gilt zwar prinzipiell auch für weitere Parameterschätzgrößen, wobei diese für inhaltlich-substanzielle Untersuchungen eher von sekundärem Interesse sein werden. Parameterverzerrungen in den Messmodellen oder in den Fehler(ko)varianzen sind zu vernachlässigen, wenn trotzdem die korrekten Wachstumsverlaufsmuster des DGP in den simulierten Stichproben identifiziert werden können. Daher werden in dieser Untersuchung vorrangig die Abweichungen der Intercepts und Slopes als Performanzkriterium berücksichtigt. Da der absolute Abweichungswert zwischen dem Mittelwert der Parameter aller simulierten Stichproben und dem Parameter des DGP nicht skaleninvariant ist, bietet sich hier die Betrachtung der relativen Abweichung an. Über die Formel

$$100 * \frac{(\overline{Parameterschaetzwert} - \text{Parameter im DGP})}{\text{Parameter im DGP}} \quad (15)$$

lässt sich der prozentuale Bias einer Parameterschätzung berechnen. Der relative Bias ist in seiner Interpretation damit skaleninvariant. Als weicher und harter Grenzwert für den relativen Bias werden üblicherweise 10 % und 5 % veranschlagt (vgl. z. B. Koo & Leite, 2014: 213; Depaoli, 2013: 197; Serang et al., 2016: 5; Kim & Kim, 2012: 446). Für Parameter mit einem absoluten Populationswert von Null, kann der relative Bias nach der Formel 15 leider nicht berechnet werden. In diesem Fall muss der absolute (raw) Bias betrachtet werden.

Coverage Probabilities. Die Coverage Probabilities (CP) geben für einen Parameter den Anteil der simulierten Stichproben an, welche in ihrem Konfidenzintervall den wahren Wert des Parameters aus dem DGP enthalten (vgl. Carsey & Harden, 2014: 92 f.). Abweichungen vom veranschlagten Konfidenzintervall (z. B. 95 %) in den Coverage Probabilities deuten auf eine Verzerrung der Schätzung der Standardfehler hin. Fällt die CP kleiner als das Konfidenzintervall (z. B. 91 %) aus, deutet dies auf zu kleine Standardfehlerschätzungen hin. Enthalten die simulierten Stichproben hingegen zu häufig den wahren Parameter des DGP, dann deutet dies auf die Schätzung von zu großen Standardfehlern hin. Bisherige Untersuchungen empfehlen, dass für eine 95 % Cover.Prob. ein Wertebereich von 0.92 bis 0.97 als akzeptabel anzusehen ist (vgl. Muthén & Muthén, 2002: 606; Kim & Kim, 2012: 446). Anhand der Coverage Probabilities kann somit zusätzlich die Effizienz

der Schätzungen bewertet werden. Allerdings hängt die CP aufgrund der Berechnungslogik immer auch mit der Verzerrung der Parameterschätzungen zusammen, wodurch eine klare Trennung zwischen «Bias» und Standardfehlerverzerrung nicht möglich ist (außer es gibt keinen Parameterbias).

Standard Deviation Method (bzw. SE-Bias). Für eine isolierte Betrachtung des Standardfehlers eignet sich die Standard Deviation Method. Hierfür wird anhand der Replikationen ein empirischer Standardfehler berechnet, welcher als Quasi-Populationswert agiert gegen welchen die Standardfehlerschätzungen der Replikationen verglichen werden können. Die Unterschiedlichkeit der Parameterschätzer der Replikationen dient dabei als Schätzung der tatsächlichen Parametervariabilität bei unendlichen vielen Stichprobenziehungen. Auch der SE-Bias kann in den relativen SE-Bias umgerechnet werden. Dabei kann die Formel 15 analog verwendet werden.

Unzulässige Lösungen (Inadmissible solutions). Da es sich bei GMMs und insbesondere bei MIGMMs um komplexe iterative Schätzverfahren handelt, können Modellschätzungen auch bei nicht klaren und unzufriedenstellenden Lösungen angelangen. So zeigte sich bereits, dass insbesondere bei vielen latenten Klassen und bei geringen Fallzahlen in Kombination mit komplexen Modellen mit negativen Varianzen oder auch mit nicht-konvergierenden Lösungen zu rechnen ist (vgl. Diallo et al., 2016: 21). Unter «unzulässigen Lösungen» sind alle Modellschätzungen in den simulierten Stichproben gemeint, welche nicht konvergieren oder Parameterschätzungen außerhalb der gültigen Wertebereiche enthalten.²⁸ In Anlehnung an Forschungspraxis von Simulationsstudien bzgl. Mixture Modellen werden Modellschätzungen dann als nicht-konvergierend behandelt, wenn kein Set von Startwerten bei einer konvergierenden Lösung angelangt (vgl. auch Diallo et al., 2016: 21 f.; Liu & Hancock, 2014: 564 ; Bauer & Curran, 2003: 347).

Diese Modelle sind in der Regel nicht interpretationsfähig und in der empirischen Praxis daher zum Großteil nutzlos bis hin zu irreführend. Da unzulässige Lösungen in der empirischen Praxis oft ignoriert oder als Hinweise für zu viele latente Klassen aufgefasst werden, können sie die identifizierte Klassenanzahl maßgeblich beeinflussen. Diese empirische Praxis ist insofern problematisch, als dass sich zeigen lässt, dass unzulässige Lösungen sowohl bei zu vielen, bei zu wenigen, aber auch bei der exakt richtigen Anzahl

²⁸Hierzu zählen negative Varianzschätzungen und Korrelationswerte über- und unterhalb der Grenze von ± 1 .

von latenten Klassen auftreten (Liu & Hancock, 2014: 564 f.) und somit weder ein Hinweis für die eine oder die andere Richtung darstellen. Vor dem Hintergrund dieser Praxis gilt es daher neben den erwähnten Zielgrößen auch unzulässige Lösungen von Modellen als Zielgröße zu untersuchen, um die empirische Forschung anzuleiten.

7.2 Konstante Modellparameter

Die konstanten Modellparameter umfassen alle Aspekte der Simulation, welche nicht variiert werden. Daher kann auch der Einfluss der konstanten Aspekte im Rahmen der MC-Studien nicht untersucht werden. Streng genommen gelten die Ergebnisse und Schlussfolgerungen von MC-Simulationsstudien prinzipiell nur für Modelle, welche die gleichen konstanten Größen aufweisen wie die Simulation selbst. Daher ist es zentral, hierbei Werte auszuwählen, welche dem Durchschnitt der in der empirischen Praxis vorzufindenden Werte entsprechen. So wird sichergestellt, dass die abgeleiteten Ergebnisse tatsächlich für die Forschungspraxis relevant sind. Im Folgenden werden die zentralen konstanten Größen der Simulation dargestellt und in ihrer jeweiligen Ausführung begründet.

Messzeitpunkte. Die Anzahl der Messzeitpunkte wird auf 5 Erhebungswellen festgelegt. Da große sozialwissenschaftliche Surveys in der Regel auf jährlichen Messungen beruhen und kleinere Erhebungen aufgrund fehlender finanzieller Ausstattung auf wenige Wellen beschränkt bleiben, sind häufigere Erhebungswellen in der Praxis selten anzutreffen. Auch vergleichbare Simulationsstudien arbeiten mit einer ähnlichen Anzahl (4–6) an Messzeitpunkten (z. B.: Diallo et al., 2016; Kim & Willson, 2014; Wirth, 2008; Serang et al., 2016). Die Anzahl der Messzeitpunkte weist im Kontext von GMMs eine ähnliche Rolle wie die Fallzahl auf. Je höher die Anzahl, desto bessere Schätzergebnisse sind zu erwarten, da mehr Informationen zu den Wachstumsverlaufsgruppen zur Verfügung stehen. So konnte auch bereits gezeigt werden, dass die Häufigkeit der Messung (Anzahl Erhebungswellen) im Rahmen von GMMs die Fallzahl bis zu einem gewissen Grad kompensieren kann, was die Qualität der Schätzergebnisse angeht (vgl. Kim & Kim, 2012: 437). Darüber hinaus zeigt die Anzahl der Messzeitpunkte in Simulationsstudien (zumindest im Kontext von GMMs) keine bedeutenden Effekte (vgl. Tofghi & Enders, 2008: 331).

Manifeste Indikatoren. Für die manifesten Indikatoren pro Messzeitpunkt wird eine geringere Anzahl festgelegt, wie in vergleichbaren Simulationsstudien. Die dort veranschlagte Indikatorenzahl von 6 bis 8 erscheint aus einer praktischen Anwendungsperspektive deutlich zu hoch. In dieser Simulation werden daher lediglich 4 Indikatoren für einen Messzeitpunkt veranschlagt, was der Forschungspraxis in der empirischen Sozialforschung näher kommen sollte und zugleich die Modellkomplexität nicht unnötig erhöht.

Missing Values. Bei der Festlegung des Datengenerierungsprozesses werden zusätzlich keine fehlenden Werte berücksichtigt, sondern nur vollständige Daten erzeugt. Auch wenn man empirisch nie von vollständigen Fällen ausgehen kann, soll hierdurch die Komplexität des Simulationsmodells handhabbar bleiben. Abgesehen von Verzerrungen durch Nonignorable Missing, überschneidet sich der Anteil der Missings (ebenso wie die Anzahl der Messzeitpunkte) mit der benötigten Fallzahl.

Wachstumsverlauf und Kovarianzen. Da der Fokus dieser Studie nicht auf der Wachstumsverlaufsform liegt, wird ein einfaches Wachstumsmuster spezifiziert, welches sich zudem nicht zwischen den latenten Gruppen unterscheidet. So wird lediglich ein linearer Wachstumsterm im DGP berücksichtigt, welcher sich lediglich in seinen Parameterausprägungen zwischen den Gruppen unterscheidet, aber nicht in der allgemeinen Form. Auch die Kovarianz zwischen den Wachstumsparametern kann gruppenspezifisch unterschiedlich modelliert werden. Allerdings wurde dieser Aspekt bereits hinreichend für GMMs untersucht²⁹ und wird somit als invariant spezifiziert.

Diachrone Fehlerkorrelationen. Auch Fehlerkovarianzen der manifesten Indikatoren werden zwischen den gleichen Variablen der fünf Erhebungszeitpunkte modelliert. Diese Form von diachronen Korrelationen stellt in Längsschnittmodellen mit Mehrindikatormessung die Regel dar und soll somit auch für den DGP berücksichtigt werden (vgl. hierzu auch Urban & Mayerl, 2014: 173 f.; Urban, 2004: 44 ff.). Obwohl die Fehlervarianzen gruppenunterschiedlich spezifiziert werden müssen (siehe Abschnitt 7.3), werden die längsschnittlichen Fehlerkovarianzen als gruppeninvariant simuliert. So mag es zwar plausibel sein, dass Indikatoren zwischen zwei Gruppen unterschiedlich gut messen, aber wieso

²⁹Das Fazit bezüglich der freien und invarianten Modellspezifikation von Wachstumsparameterkovarianzen lautet, dass freie Schätzungen für jede latente Klasse immer vorzuziehen sind (vgl. Diallo et al., 2016: 20 f.). Tatsächlich gehört eine korrekt spezifizierte Mittel- und Kovarianzstruktur zu den Grundvoraussetzungen von GMMs (vgl. Bauer, 2007: 767 f.).

die bei der Messung entstehenden Fehler für verschiedene Gruppen unterschiedlich stark miteinander assoziiert sein sollen, ist weniger eindeutig. Die zusätzliche Komplexität dieses Aspektes wird bei dieser Simulation folglich ausgeschlossen. Simulationsstudien weisen zusätzlich darauf hin, dass die Fehlspezifikation von zeitlichen Fehlerassoziationen (zumindest für GMMs) zwar Parameterschätzungen verzerren kann, aber bei der korrekten Klassenidentifikation wenig ins Gewicht fällt (vgl. Koo & Leite, 2014: 223 f.).

Anzahl latenter Klassen. Die Anzahl der tatsächlich in den Daten vorhandenen latenten Klassen wird auf drei fixiert. Da die reine Anzahl der Klassen keine Auswirkungen auf das Schätzverfahren hat, wird hierdurch die Gültigkeit und Übertragbarkeit der Ergebnisse nicht eingeschränkt. Nur absolute und relative Klassengrößen könnten einen Effekt haben, welcher allerdings innerhalb dieser Studie bereits über die variierenden Simulationsgrößen (Fallzahl und Mixing Proportions) abgebildet werden. Daher ist es lediglich nötig einen möglichst repräsentativen Wert aus der empirischen Praxis festzulegen.

Enders & Tofighi, 2008 kommen anhand einer Literatursichtung zu dem Schluss, dass sich die typische Anzahl von extrahierten Klassen in der empirischen Praxis zwischen 2 und 4 bewegt, wobei eine Extraktion von mehr als 4 Klassen sehr selten vorkam (vgl. Enders & Tofighi, 2008: 80). Auch Kim & Kim (2012) kommt zu dem Ergebnis, dass sich die extrahierten Klassen in der GMM-Literatur im Bereich von 2 bis 7 mit Modi von 3 und 5 bewegen (vgl. Kim & Kim, 2012: 464 ff.). Auch die große Mehrzahl der für diese Studie gesichteten GMM-MC-Untersuchungen bewegte sich im Bereich von 2 bis 4 simulierten latenten Klassen, wenn die Klassenzahl nicht systematisch in der Studie variiert wurde. Darüber hinaus werden die hier festgelegten drei latente Klassen auch in vielen vorangegangenen GMM-Simulationsstudien als Ausgangspunkt gewählt, wodurch die Anschlussfähigkeit, hinsichtlich deren Ergebnisse erhöht wird (siehe z. B.: Depaoli, 2013; Diallo et al., 2016; Enders & Tofighi, 2008; Peugh & Fan, 2012).

Indikatorskalierung und Modellschätzung. Die manifesten Indikatoren des Simulationsmodells werden als metrisch skaliert konstruiert (im DGP) und im Rahmen des Schätzmodells für die einzelnen Replikationen auch als metrisch angenommen. Alle Modellreplikationen der Simulationsstudie werden innerhalb der Statistiksoftware Mplus mit dem MLR-Schätzer berechnet. Der MLR-Schätzer in Mplus stellt einen Maximum-Likelihood-Parameter-Schätzung mit Standardfehlern und einer χ^2 -Teststatistik dar, welche robust

gegenüber Nicht-Normalität ist. Die MLR-Standardfehler werden mit Hilfe eines Sandwich-Schätzers berechnet. Die MLR χ^2 -Teststatistik ist asymptotisch äquivalent zur Yuan-Bentler T2*-Teststatistik (vgl. Muthén & Muthén, 2015: 608). Das Schätzverfahren muss somit ebenfalls als ein eine konstante Größe der MC-Studie verstanden werden. Der MLR-Schätzer wird breit angewendet und stellt den De-facto-Standard für GM-Modelle dar.

7.3 Variierte Modellparameter

Diese Variablen repräsentieren die Variationsstufen der Simulation. Ihr Einfluss auf die Zielgrößen soll in dieser Untersuchung erörtert werden. Die Schwierigkeit besteht darin, geeignete Ausprägungskategorien für die unabhängigen Variablen festzulegen. Eine möglichst breite und differenzierte Skala ist zwar wünschenswert, treibt aber die Laufzeit der Simulationen in die Höhe, da jede einzelne Variablenausprägung in Kombination mit den anderen unabhängigen Variablen und deren einzelnen Ausprägungen eine Konfiguration darstellt, welche zu simulieren ist. Die Ausprägungen der unabhängigen Variablen sollten sich dabei an der Praxis in der empirischen Forschung orientieren, damit die Ergebnisse und Schlüsse aus der Simulation die empirische Forschung unterstützen und anleiten können.

7.3.1 Manipulierte Messinvarianzen

Verletzungen der Messinvarianz werden in dieser Untersuchung sowohl hinsichtlich der Item-Intercepts als auch hinsichtlich der Faktorladungen modelliert, wodurch eine Verletzung der metrischen und der skalaren Invarianz hergestellt wird. Zusätzlich wird nach zeitlicher bzw. längsschnittlicher und nach gruppenspezifischer Nicht-Invarianz unterschieden. Zur Modellierung der längsschnittlichen Invarianz werden jeweils 2 der 4 Messindikatoren über die Zeit manipuliert. Für die gruppenspezifische Nicht-Invarianz werden hingegen jeweils 3 der 4 Messindikatoren hinsichtlich Intercepts und Faktorladungen manipuliert. Dabei werden bei den gruppenspezifischen Verletzungen der Messinvarianz zwei der drei generierten latenten Klassen im Vergleich zur Referenzgruppe verzerrt.

Referenzmodell. Als Referenz wird in der ersten Ausprägung keine Verletzung der Messinvarianz angenommen. Sowohl Item-Intercepts als auch Faktorladungen bleiben über die Zeit gleich und unterscheiden sich nicht zwischen den Gruppen. Dies entspricht der üblichen Annahme von metrischer und skalarer Invarianz, welche bei der Modellierung

von GMMs vorausgesetzt wird. Für die Modellschätzungen dieser Konfiguration werden folglich keine Verzerrungen (über das zufällige Maß hinaus) angenommen. Als Referenzwerte werden die Faktorladungen auf den 1.0 festgesetzt und die Item-Intercepts jeweils auf den Wert 0.0 fixiert.

Längsschnittliche Nicht-Invarianz. Für die rein längsschnittliche Nicht-Invarianz wird zwischen zwei Konfigurationen unterschieden. In der ersten Konfiguration wird lediglich ein zeitlicher Wandel in den Intercepts angenommen, wobei die Faktorladungen identisch bleiben. Die Intercepts der zwei manipulierten Items werden dabei pro Erhebungswelle um den Wert von 0.1 verringert, was einer Gesamtveränderung von -0.4 entspricht. Die Größenordnung dieser Trendveränderung entspricht vorangegangenen Simulationsstudien, welche Item-Intercepts systematisch variiert haben (vgl. z. B.: Wirth, 2008; Olivera-Aguilar, 2013; Zhang, 2016). Nach Olivera-Aguilar (2013) ist diese Veränderung in etwa als eine mittlere Effektstärke zu quantifizieren (vgl. Olivera-Aguilar, 2013: 56).

In der zweiten Konfiguration werden sowohl Intercepts als auch Faktorladungen verzerrt. Für die Faktorladungen wird ebenfalls ein Trend in der Veränderung angenommen. So wird von jedem der vier folgenden Messzeitpunkte der Wert 0.08 von der vorangegangenen Faktorladung des jeweiligen Items abgezogen (über alle Messzeitpunkte folglich -0.32). Auch die Größenordnung dieser Gesamtänderung orientiert sich an bisherigen Simulationsstudien (vgl. z. B.: Kim & Willson, 2014; Wirth, 2008; Olivera-Aguilar, 2013; Zhang, 2016). Diese Verzerrung ist eher als stark einzuordnen (vgl. Olivera-Aguilar, 2013: 52 f.). Die gleichförmigen Veränderungstrends der Intercepts und Faktorladungen über die Zeit wurden hier bewusst in Übereinstimmung mit Wirth (2008) gewählt. Alternative Modellierungen hätten eine zufällige (z. B. Leite, 2007) oder wechselnde/vermischte Veränderung wählen können. Allerdings kann argumentiert werden, dass zufällige und wechselnde Veränderungen aus einer empirischen Perspektive weniger realistisch erscheinen, wie eine allmähliche systematische zeitliche Veränderung eines latenten Konstrukts über die verschiedenen Erhebungswellen (vgl. Wirth, 2008: 110). So erscheint es anschlussfähiger für die empirische Praxis einen allmählichen Wandel der Messung zu simulieren als eine zufällige Veränderung oder eine Hin und Her wechselnde Bedeutung des Messkonstrukts.

Gruppenspezifische Nicht-Invarianz. Auch für die gruppenspezifische Nicht-Invarianz wird zwischen zwei Konfigurationen unterschieden. In der ersten Konfiguration wird ledig-

lich ein Gruppenunterschied in den Item-Intercepts modelliert. In der zweiten Konfiguration wird ein Unterschied in den Item-Intercepts und in den Faktorladungen angenommen. Die Größe der Gruppenunterschiede wird an Kim & Willson (2014) angelehnt, welche die Item-Intercepts um die Beträge 0.25 und 0.5 sowie die Faktorladungen um die Beträge 0.15 und 0.3 verzerrt haben (vgl. Kim & Willson, 2014: 413). Die Stärke dieser Verzerrungen gleicht den Gesamtbeträgen der längsschnittlichen Nicht-Invarianzen, wodurch beide Konfigurationen einen ähnlich starken Einfluss ausüben können (oder zumindest hinsichtlich der Größenordnung der Manipulation vergleichbar sind). Da drei latente Gruppen modelliert werden, können auch unterschiedliche Gruppendifferenzen umgesetzt werden, welche sich in Richtung und Stärke unterscheiden. Es werden jeweils 3 der 4 Messindikatoren manipuliert. Wenn Gruppe 3 die Referenz darstellt, wird für Gruppe 1 der Intercept um -0.5 und die Faktorladung um -0.3 verzerrt. Für Gruppe 2 wird hingegen der Intercept um +0.25 und die Faktorladung um +0.15 verzerrt. Dadurch wird sowohl die negative und die positive Effektrichtung abgebildet als auch zwei unterschiedlich starke Grade an Verzerrung. Auch die Betrachtung der kombinierten Messinvarianzverletzungen wird dadurch bereichert.

Kombinierte Nicht-Invarianz. Für die kombinierte Nicht-Invarianz werden die längsschnittlichen und die gruppenspezifischen Verzerrungen zusammengefasst. Dafür werden die Messverzerrungen der Intercepts und der Faktorladungen aus den vorangegangenen Ausprägungen zusammengekommen, indem die gruppenspezifischen und die längsschnittlichen Verzerrungen addiert werden. Die gruppenspezifischen Nicht-Invarianzen enthalten sowohl eine latente Klasse, welche der Trendveränderungen der längsschnittlichen Nicht-Invarianz entspricht, als auch eine Klasse, welche der Trendveränderung entgegensteht. Dadurch kann untersucht werden, inwieweit sich die Messverzerrungen verstärken, aber auch inwieweit sie gruppenspezifische und längsschnittliche Invarianzverletzungen kompensieren können.

So werden insgesamt sechs verschiedene Kategorien hinsichtlich der Verletzung der Messinvarianz unterschieden:

1. INV: Messinvarianz
2. LNI: längssch. Nicht-Invarianz der Intercepts
3. LNIF: längssch. Nicht-Invarianz der Intercepts u. Faktorladungen

4. GNI: gruppenspez. Nicht-Invarianz der Intercepts
5. GNIF: gruppenspez. Nicht-Invarianz der Intercepts u. Faktorladungen
6. LGNIF: kombinierte Nicht-Invarianz der Intercepts u. Faktorladungen

Sicherlich wäre auch ein gänzlich faktorielles Untersuchungsdesign möglich gewesen, in welchem jede einzelne Ausprägung der Verzerrung mit jeder möglichen Kombination überprüft wird. Allerdings hätte dies die Zahl der Kombinationen sehr stark in die Höhe getrieben, wodurch der Aufwand für die Simulation zu hoch geworden wäre. Zudem hätte dies auch bedeutet, dass Datensituationen berücksichtigt werden, welche aus einer praktischen Perspektive keinerlei Relevanz für die Empirie haben. Hierbei ist z. B. an eine unterschiedliche konzeptionelle Bedeutung der Messkonstrukte bei einer identischen verwendeten Messskala zu denken, wie sie bei nicht-invarianten Faktorladungen in Kombination mit identischen manifesten Intercepts auftreten würde. Die hier getroffene Auswahl der Variationsstufen soll somit nur die wichtigsten Konfigurationen abbilden, ohne alle prinzipiell möglichen Ausprägungen der Nicht-Invarianz miteinander zu kreuzen.

7.3.2 Mixing Proportions

Die prozentuale Verteilung der Anteilsverhältnisse der latenten Klassen wird unter drei Bedingungen überprüft. (1) Vollständige Gleichverteilung (bei drei Klassen 33 %, 33 %, 33 %), (2) moderate Ungleichverteilung (17 %, 33 %, 50 %; in Anlehnung an Diallo et al. (2016)) und (3) eine Ungleichverteilung mit einer sehr kleinen Klasse (5 %, 45 %, 50 %). Bisherige Simulationsergebnisse sprechen dafür, dass moderat ungleichverteilte Klassenanteile im Vergleich zu einer völligen Gleichverteilung noch keinen Effekt haben sollten (vgl. hierzu Abschnitt 3.2.2 auf Seite 38). Erst bei extrem kleinen latenten Gruppen sind Probleme hinsichtlich der Modellperformanz zu erwarten. Die Ungleichverteilung mit der kleinen Klasse von 5 % ist hier bewusst so gewählt, dass sie die Grenze der aktuell noch als interpretationswürdig empfohlenen Klassengrößen berührt (vgl. hierzu Wickrama, 2016: 215 ff.). Auch wenn bei einer moderaten Ungleichverteilung noch nicht von verzerrenden Wirkungen auszugehen ist, besteht weiterhin die Möglichkeit, dass in Kombination mit anderen Simulationsmerkmalen in der Interaktion dennoch Effekte auf die Modellperformanz auftreten. Bisherige Ergebnisse zu ungleichverteilten Anteilswerten werden somit für den Kontext von MIGMMS auf eine zusätzliche Bewährungsprobe gestellt.

7.3.3 Fallzahl

In Anlehnung an Diallo et al. (2016) werden für die Fallzahl die Ausprägungen 300, 700 und 1300 gewählt. Wie in den meisten Simulationsstudien, in welchen die Fallzahl nicht der primäre Fokus der Untersuchung ist, wird damit eine kleine, eine mittlere und eine große Fallzahl ausgewählt. Auch Tofighi & Enders (2008) wählten ähnliche Fallzahlen nach einer umfangreicheren Sichtung der inhaltlich-substantiellen Anwendungen von GM-Modellierungen (vgl. Tofighi & Enders, 2008: 325). Vorangegangene Untersuchungen haben gezeigt, dass die Stichprobengröße für die Identifikation der korrekten Anzahl an latenten Klassen entscheidend sein kann (siehe hierzu ausführlich Abschnitt 3.2.4 auf Seite 40). Generell ist anzunehmen, dass höhere Fallzahlen die Chance erhöhen anteilsmäßig kleine, aber inhaltlich relevante, latente Klassen zu finden. Darüber hinaus ist davon auszugehen, dass höhere Fallzahl zu weniger unzulässigen Lösungen führt, da mit einer besseren Modellkonvergenz und mit weniger unzulässigen Parameterschätzungen zu rechnen ist (vgl. Berlin et al., 2014: 14).

7.3.4 Klassenseparation

Eine geringe Klassenseparation ist mit einer Vielzahl von Problemen hinsichtlich der Modellperformanz verbunden (siehe hierzu ausführlich Abschnitt 3.2.2 auf Seite 38). In Kombination mit fehlender Messinvarianz ist zu erwarten, dass die Klassenidentifikation, Parameterverzerrungen sowie unzulässige Lösungen weiter zunehmen. Im Gegensatz dazu ist davon auszugehen, dass die Probleme von falschen Invarianzannahmen bei einer hohen Klassenseparation zu weniger Problemen hinsichtlich der Modellperformanz führen. Ähnlich ausgeprägte Effektrichtungen sind auch für die Fallzahl und für die Anteilsverhältnisse zu erwarten. Für diese Simulation wird zwischen einer geringen, einer mittleren und einer extremen Klassenseparation unterschieden (Mahalanobis Distance (MD) 0.8, 1.5, 4.0). Die Distanz ist hierbei jeweils auf die benachbarten Klassen bezogen. D. h. für die Simulationsausprägung «0.8» wäre die latente Klasse 1 von Klasse 2 um 0.8 MD entfernt (Klasse 2 von Klasse 3 ebenfalls um 0.8 MD) wohingegen Klasse 3 von Klasse 1 um 1.6 MD ($0.8 + 0.8$) Einheiten entfernt wäre. Die hier ausgewählten Ausprägungen der Klassenseparation bewegen sich im Rahmen von bisherigen Simulationsanalysen, welche die Klassendistanz systematisch variieren (vgl. z. B.: Depaoli, 2013; Liu & Hancock, 2014; Lubke & Neale, 2006; Peugh & Fan, 2012; Zhang, 2016). Die extreme Ausprägung

von 4.0 MD weicht von den üblichen Simulationsdesigns leicht ab. Allerdings dient sie in dieser Simulation als Quasi-Nullausprägung, da Klassenüberschneidungen bei dieser klaren Trennung keine Auswirkungen haben sollten.³⁰ Die Unterschiede der latenten Klassen werden im DGP über die gemeinsame Veränderung der Slopes und Intercepts hergestellt. Klassenunterschiede, welche nur auf einer Variation der Slopes oder nur der Intercepts beruhen, werden nicht berücksichtigt. Liu & Hancock (2014) liefern für dieses Vorgehen überzeugende Argumente. Sie stellen heraus, dass zum einen gleiche Startniveaus des Wachstums in der Praxis selten vorkommen werden und zum anderen gleiche Wachstumsraten mit unterschiedlichen Startniveaus ebenso gut über LGC-Modelle mit einer größeren Interceptvarianz abgebildet werden und daher keinen genuinen Anwendungsfall von GMMs darstellen (vgl. Liu & Hancock, 2014: 561).

7.3.5 Zusammenfassung bzgl. variierender Größen

Alle hier aufgezeigten variierenden Simulationsgrößen werden in Form eines vollständigen faktoriellen Designs in dieser MC-Studie berücksichtigt. Daher sind sowohl die Haupteffekte der einzelnen Größen identifizierbar wie auch die Interaktionseffekte zwischen den Variablengrößen. Durch die Anzahl der unabhängigen Größen und deren Ausprägungen gelangt diese Studie auf 162 ($6 \cdot 3 \cdot 3 \cdot 3$) Simulationskonfigurationen. Bei 300 Replikationen pro Simulationskonfiguration, jeweils 4 Modellschätzungen für die unterstellte latente Klassenanzahl und 200 randomisierten Startwerten pro Modell um lokale Maxima zu vermeiden, gelangt man bei insgesamt 38 888 000 einzelnen Modellschätzungen³¹ an. Dies verdeutlicht, wieso die Parallelrechner des BwUniCluster zur Durchführung dieser Untersuchung essenziell sind und wieso nicht weitere Konfigurationen bzw. Variablenausprägungen berücksichtigt werden können.

Zur Untersuchung der Klassenidentifikation nach dem BLRT muss aufgrund des extrem erhöhten Rechenaufwandes des BLRT die Anzahl der Replikationen auf 200 heruntergesetzt werden und die Anzahl der randomisierten Startwertesets auf 120 verringert werden.³² Dabei handelt es sich folglich um eine eigene kleinere Simulationsstudie, welche nicht auf Grundlage der gleichen Grundgesamtheit operiert, wie die restlichen Berech-

³⁰Allerdings fällt diese Ausprägung der Klassendistanz auch nicht völlig aus dem Rahmen. Zhang (2016) verwendet sogar die Ausprägung 5.0 MD, um eine eindeutige Klassenteilung herzustellen.

³¹Ohne explizite Berücksichtigung der randomisierten Startwerte gelangt man bei 194 400 Modellschätzungen an.

³²Die Modellberechnungen für 88 von 648 Klassenmodellen hat in Testläufen die maximal gestattete Laufzeit von 6 Tagen auf dem BwUniCluster überschritten und konnten somit nicht mit der angestrebten Replikationszahl ausgewertet werden.

nungen. Da allerdings der gleiche Seed gewählt wurde, gleichen sich die Ergebnisse der ersten 200 Replikationen für die Hauptsimulationsstudie und für die BLRT-Simulation, was einen gewissen Grad an Vergleichbarkeit aufrecht erhält. Die Replikationen der BLRT-Simulation stellen damit eine Teilmenge der Replikationen der Hauptstudie dar. Folglich werden für die Auswertung nach dem BLRT lediglich 129 600 ($162 \cdot 200 \cdot 4$) Modellschätzungen durchgeführt.

Variable	Variierte Simulationsgrößen	
	Anzahl Kategorien	Ausprägungen
Messinvarianz	6	INV (invar=1) LNI (invar=2) LNIF (invar=3) GNI (invar=4) GNIF (invar=5) LGNIF (invar=6)
Mixing Proportions	3	33 %, 33 %, 33 % (mixing=1) 17 %, 33 %, 40 % (mixing=2) 5 %, 45 %, 50 % (mixing=3)
Fallzahl	3	300 Fälle (nobs=1) 700 Fälle (nobs=2) 1300 Fälle (nobs=3)
Klassenseparation	3	0.8 MD (sep=1) 1.5 MD (sep=2) 4.0 MD (sep=3)
	Konstante Simulationsgrößen	
	Anzahl Kategorien	Ausprägung
Messzeitpunkte	1	5 Erhebungswellen
Anzahl manifeste Indikatoren	1	4 Items pro Welle
Missing Values	1	keine fehlenden Werte
Wachstumsverlaufmuster	1	lineare Wachstumsfkt.
(Ko-)Varianz Wachstumsfaktoren	1	gruppeninvariant
Diachrone Fehlerkorrelationen	1	gruppeninvariant
Anzahl latenter Klassen	1	3 latente Klassen
Anzahl Konfigurationen	162	

Tabelle 1: Übersicht konstanter und varianter Simulationsgrößen

8 Simulationskonfiguration und Populationsparameter

Die vorangegangenen Kapitel dienten dazu, die Forschungslücken hinsichtlich der Leistungsfähigkeit von MIGMMs darzulegen und die Bearbeitungsstrategie zur Schließung dieser Lücken zu illustrieren. Hierzu wurden Arbeitshypothesen entwickelt sowie die Zielgrößen und variablen Simulationsparameter gerechtfertigt, welche die Bearbeitung der Hypothesen ermöglichen.

In den folgenden Kapiteln sollen die Ergebnisse der Simulationsstudie hinsichtlich der Identifikation der latenten Klassen dargestellt und bewertet werden. Dazu gilt es zunächst die Simulation in allen Einzelheiten darzustellen. Der erste Abschnitt 8 beschäftigt sich daher mit den bisher noch nicht gänzlich dargestellten fixen Parametern der Simulationskonfigurationen. Da deren Festlegung prinzipiell kontingent ist, gilt es die einzelnen Werte zu rechtfertigen. Zugleich werden dabei die bereits erwähnten variablen Parameter überblicksartig dargestellt. Dies soll die Nachvollziehbarkeit, Replikation und potentielle Erweiterung der Simulationsergebnisse vereinfachen. Hierzu gehört auch die Dokumentation der praktischen EDV-technischen Durchführung der Berechnungen für die Simulation. Dabei wird nicht nur auf das Datenmanagement, sondern auch auf Sonderthemen wie die Probleme der Zufallsziehung und das Label-Switching im Rahmen von Mixture-Simulationsmodellen eingegangen.

Die weiteren Abschnitte 9 bis 11 befassen sich dann schließlich mit der Ergebnisdokumentation der Simulationen und der eigentlichen Bearbeitung der in Abschnitt 6 aufgestellten Thesen. Darüber hinaus sollen hieraus Handlungsempfehlungen für die empirische Forschungspraxis erarbeitet werden. In dieser ersten Ergebnisanalyse steht die korrekte Identifikation von latenten Klassen im Rahmen von MIGMMs anhand der IC-Indizes AIC, BIC und aBIC im Mittelpunkt, was durch eine Sonderauswertung zum BLRT ergänzt wird. Hierbei wird auch zusätzlich auf deren Konditionalität bzgl. der weiteren variierten Simulationsgrößen eingegangen.

Folglich sollen im Anschluss an die vorangegangenen Ausführungen zu den zentralen variablen und konstanten Parametern der Simulationsstudie die Einzelheiten der Simulationskonfigurationen zunächst detailliert aufgeschlüsselt werden. Erst eine mathematisch exakte Repräsentation der Untersuchung, welche den Anwendungskontext der Analyse beschreibt, erlaubt eine Entscheidung über die Relevanz und Nützlichkeit der Simulation. Zu diesem Zweck wird zunächst auf implizite Modellannahmen für MI/SO-

Wachstumsmodelle eingegangen um im Anschluss hieran die festgelegten Populationswerte zu rechtfertigen und darzustellen. Nach der Beschreibung des DGP über die Populationswerte gilt es noch das Schätzmodell darzustellen, welches anhand der generierten Daten diejenigen Modellparameter schätzt, welche wiederum später anhand ihrer Übereinstimmung mit den Populationswerten zu evaluieren sind.

8.1 Modellannahmen von MI/SO-Wachstumsmodellen

Konfigurale, metrische und skalare Messinvarianz muss im Rahmen von MIGMMs überprüft werden um unverzerrte Wachstumsverläufe nachweisen zu können. Strengere Formen der Messinvarianz können allerdings im Rahmen von MIGMMs zu messtechnischen Problemen führen und stellen oft keine theoretisch angemessene Modellierung dar (vgl. Wirth, 2008: 30 ff.). Da diese Probleme unmittelbar mit der Festlegung der Fehlerterme des DGP dieser Simulation assoziiert sind, soll hier kurz darauf eingegangen werden. Die Varianz eines manifesten Indikators y_{jt} ³³ ist im Rahmen eines MIGMMS (bzw. auch bei SO-LGC-Modellen) definiert über:

$$\text{VAR}(y_{jt}) = \lambda_{jt}^2 * \text{VAR}(\eta_t) + \epsilon_{jt} \quad (16)$$

, wobei η_t für das latente Konstrukt zum Messzeitpunkt t steht, λ_{jt} den Pfadkoeffizienten vom latenten Konstrukt zum jeweiligen manifesten Indikator j benennt und ϵ_{jt} den Fehlerterm der manifesten Variablen bezeichnet (vgl. Wirth, 2008: 30). Die Varianz des latenten Konstrukts η_t ist wiederum definiert über:

$$\text{VAR}(\eta_t) = \gamma_I^2 * \text{VAR}(I) + \gamma_{S,t}^2 * \text{VAR}(S) + 2(\gamma_I * \gamma_{S,t} * \text{KOVAR}(I, S)) + \zeta_t \quad (17)$$

(vgl. Wirth, 2008: 30), wobei γ für die Pfadkoeffizienten der Wachstumsparameter (I für Intercept und S für Slope) auf die latenten Konstrukte steht und ζ_t den Fehlerterm der latenten Konstrukte beschreibt. Da Formel 17 den Term $\gamma_{S,t}$ enthält, wächst die Varianz von η zu jeder Erhebungswelle an, wenn der Fehlerterm ζ konstant bleibt. Jede weitere Welle ist mit einem höheren Wert für den Pfad ($\gamma_{S,t}$) vom Slope zu den latenten Messmodellen parametrisiert. Verbindet man diese Schlussfolgerung mit Formel 16, erkennt man,

³³Der Laufindex j steht für die Anzahl an manifesten Indikatoren pro Welle. Der Laufindex t steht hingegen für die jeweilige Erhebungswelle.

dass der gleiche Mechanismus bei der Varianz der manifesten Indikatoren für verschiedene Erhebungswellen greift. Bei einem konstanten Fehlerterm ϵ nimmt die Varianz, bedingt durch den Formelbestandteil $\text{VAR}(\eta_t)$, über die Zeit automatisch zu.

Nimmt man folglich strengere Formen der Messinvarianz an (z. B. der Faktorfehlervarianzen und/oder der Fehlerterme der manifesten Indikatoren), so hat dies Konsequenzen für die Varianzen der latenten und manifesten Indikatoren. Zugleich sind hierdurch allerdings auch die erklärten Varianzanteile (z. B. der R^2 -Wert) über die Zeit betroffen. Dies gilt sowohl für die latenten Messfaktoren als auch für die manifesten Messungen. Da der erklärte Varianzteil der latenten Messmodelle und der manifesten Indikatoren über

$$\begin{aligned} R_{y_{jt}}^2 &= 1 - \frac{\epsilon_{kt}}{\text{VAR}(y_{jt})} \\ R_{\eta_t}^2 &= 1 - \frac{\zeta_t}{\text{VAR}(\eta_t)} \end{aligned} \tag{18}$$

gegeben ist (vgl. Wirth, 2008: 31 f.), folgt aus einer Vergrößerung der Varianz von y_{jt} und η_t eine Steigerung des R^2 -Wertes. Falls die jeweiligen Fehlerterme konstant bleiben würden, würde der erklärte Anteil der Varianz bei steigenden Erhebungswellen gegen 100 % konvergieren.³⁴

Für die hier durchgeführte Simulation bedeutet dies, dass man die Fehlerterme über die Zeit hinweg anpassen muss, damit der erklärte Varianzanteil der Messfaktoren und der manifesten Variablen konstant bleibt. Die entsprechenden Modifikationen an den Fehlervarianzwerten des DGP sind in Abschnitt 8.2.2 dargestellt.

8.2 Ausgangsmodell

Die bisherigen Ausführungen haben sich auf die zentralen variierten Größen der Simulation konzentriert (vgl. das vorangegangene Kapitel 7.3 zur Rechtfertigung der Simulationsgrößen). Vereinzelt wurde auch schon bereits auf Modellgrößen eingegangen, welche im Rahmen der Simulationsstudie nicht variiert werden (z. B. Kapitel 7.2). Trotz dessen wurden dabei nicht alle Modellgrößen abschließend festgelegt. Im Weiteren wird zunächst das Ausgangsmodell der Simulationsstudie dargestellt. Dabei wird insbesondere auf die Themen Wachstumsverlauf, Fehler(ko)varianzen, Faktorladungen und Intercept eingegangen. Im Rahmen dieser Ausführungen werden auch die Auswirkungen der Varia-

³⁴Diese Beziehung gilt nur dann nicht, wenn die Slopeparameter eine Varianz und Kovarianz von 0 aufweisen (vgl. Wirth, 2008: 32).

tion der unabhängigen Größen der Simulationstudie auf die Modellparameter dargestellt. Auf Grundlage der weiteren Ausführungen ist somit eine vollständige Replikation der Modellkonfigurationen und Ergebnisse möglich.

8.2.1 Populationswerte Wachstumsverlauf

Wie bereits dargestellt, wird für das Wachstumsverlaufmuster eine möglichst einfache Form gewählt. So wird für das Wachstum ein linearer Verlauf angenommen. Dieser Verlauf wird zwar in der numerischen Ausprägung zwischen den generierten (latenten) Gruppen variiert, aber nicht in seiner linearen Grundform. Die jeweiligen Mean-Werte können aus Tabelle 2 entnommen werden.

Tabelle 2: Populationswerte der Wachstumsparameter der Simulation

Populationswerte des Wachstumsverlaufs									
	Klasse 1			Klasse 2			Klasse 3		
Klassendistanz MD:	0.8	1.5	4.0	0.8	1.5	4.0	0.8	1.5	4.0
Konstante Klassen-Parameter									
Var(Interceptfaktor)	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
Var(Slopefaktor)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Covar(Intercept, Slope)	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Variable Klassen-Parameter									
Mean(Interceptfaktor)	0.31	0.50	1.00	0.20	0.30	0.50	0.00	0.00	0.00
Mean(Slopefaktor)	0.412	0.695	1.70	0.252	0.395	0.90	0.10	0.10	0.10

Info: Zur Herstellung der Klassendistanzen wurden die Means der Intercept- und des Slopefaktoren innerhalb der Klassen manipuliert. Die Kovarianzstruktur der Wachstumsparameter wurde als invariant über die Klassen spezifiziert.

Der Mean des Intercept- sowie des Slope-Faktors wurden dabei so zwischen den Klassen variiert, dass die Mahalanobis Distance (MD) zwischen den Klassen die Werte 0.8, 1.5 und 4.0 annimmt. Die sich daraus ergebenden durchschnittlichen Wachstumsverlaufsformen sind in Abbildung 5 (Seite 77) für die jeweils drei latenten Klassen visualisiert. Die Kovarianzstruktur bleibt hingegen über die Klassen hinweg konstant. Auch diese Werte können aus Tabelle 2 entnommen werden. Für die Varianz des Intercepts wurde in jeder Klasse der Wert 0.25 festgelegt, wohingegen die Varianz des linearen Slope-Parameters auf 0.04 fixiert wurde. Damit entspricht das Verhältnis von Intercept- und Slope-Varianz in etwa den Empfehlungen von Muthén & Muthén (2002) von 5:1 (vgl. Muthén & Muthén, 2002: 604). Die Kovarianz wurde auf einen positiven Wert von 0.02 festgelegt. Höhere

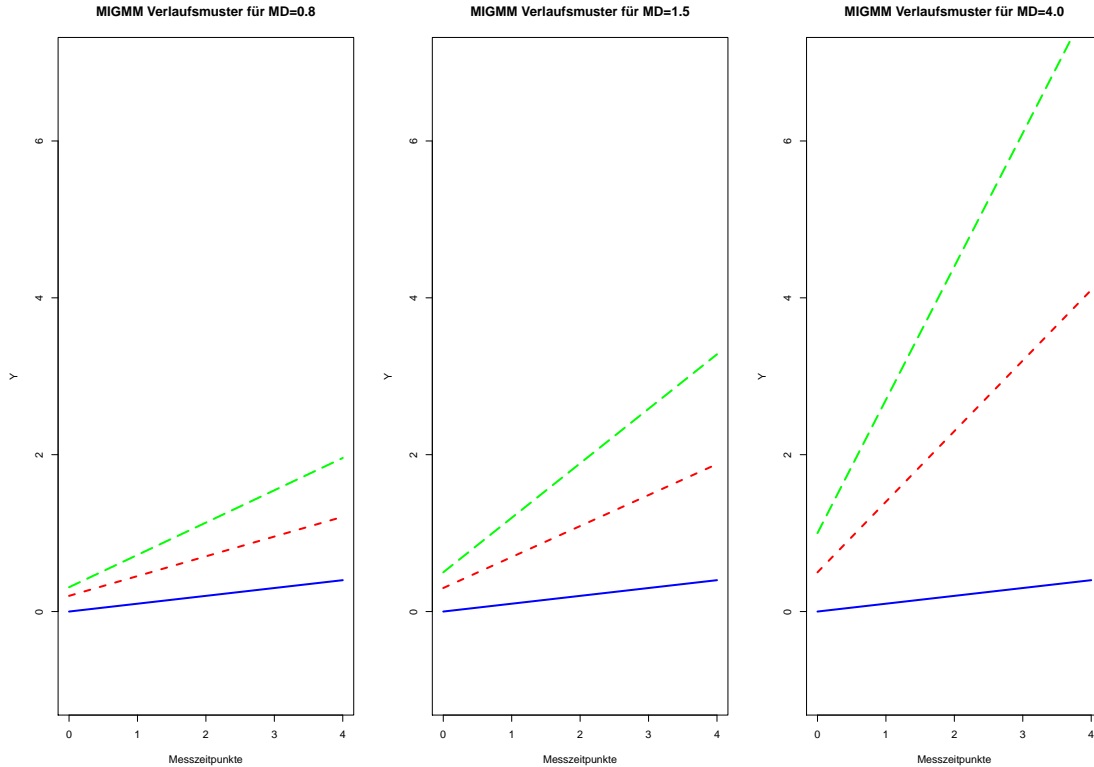


Abbildung 5: Durchschnittliches Verlaufsmuster der latenten Wachstumsklassen nach den Abstufungen der Klassendistanz

Intercept-Werte sind somit mit steileren Steigungen assoziiert. Dies entspricht einer Korrelation von 0.2 zwischen Intercept- und Slope-Faktor in jeder latenten Klasse. Auch wenn andere Autoren hier Korrelationen von bis zu 0.4 vorschlagen (z. B.: Wirth, 2008; Leite, 2007), wurde in Übereinstimmung mit Zhang (2016) der niedrigere Wert von $r = 0.2$ veranschlagt, um die Anschlussfähigkeit und Übertragbarkeit der Ergebnisse im Themenbereich von MIGMMS zu vereinfachen.

8.2.2 Populationswerte Fehlervarianzen

Da die vom Modell unterstellte Kovarianzmatrix eines SOGMMs bei konstanten Fehlervarianzen je nach Erhebungswelle andere Erklärungsanteile implizieren würde (siehe hierzu Abschnitt 8.1), müssen die Fehlerterme dynamisch an die Erhebungswellen angepasst werden. Die hierfür veranschlagten Nachskalierungen der Fehlerterme in Abhängigkeit von den Erhebungswellen können Tabelle 3 entnommen werden.

Durch diese Anpassungen werden zu jeder Welle ca. der gleiche Varianzanteil der latenten Faktoren als auch der ungefähr gleiche Varianzanteil der manifesten Indikatoren erklärt. Die Fehlervarianzen der latenten Faktoren sind ebenfalls an die Werte aus der

Tabelle 3: Skalierte Populationswerte der Fehlervarianzen der latenten Faktoren (ζ) und manifesten Variablen (ϵ).

Populationswerte: Fehlervarianzen (ϵ und ζ)					
Fehlerterme von	Wellen i				
	1	2	3	4	5
Fehlerterme manifeste Indikatoren (ϵ)					
y_{1i}	0.20	0.27	0.40	0.59	0.84
y_{2i}	0.20	0.27	0.40	0.59	0.84
y_{3i}	0.20	0.27	0.40	0.59	0.84
y_{4i}	0.20	0.27	0.40	0.59	0.84
Fehlerterme latente Variablen (ζ)					
F_1	0.06				
F_2	0.08				
F_3	0.12				
F_4	0.18				
F_5	0.24				

Info: Mit der wellenspezifischen Anpassung/Skalierung von ζ entspricht der erklärte Varianzanteil der latenten Konstrukte F_x zu jeder Welle ca. 80 %. Mit der Anpassung/Skalierung von ϵ werden die Indikatoren zu jeder Welle ca. zu 60 % statistisch erklärt.

Simulation von Zhang (2016) angelehnt um die Vergleichbarkeit zu erhöhen (vgl. Zhang, 2016: 55).

8.2.3 Tabellarische Darstellung der Invarianzen von Intercepts und Faktorladungen

Der Effekt von verschiedenen Formen der Messinvarianz auf die Performanz von MIGMMs ist von zentralem Interesse in dieser Untersuchung. Die hier untersuchten gruppen- und zeitspezifischen Messinvarianzen werden über die Manipulation der Intercepts und der Faktorladungen der manifesten Indikatoren hergestellt. Tabelle 5 und Tabelle 4 enthalten die Werteausprägungen für die verschiedenen hier untersuchten Invarianzstufen. Veränderungen der Intercepts und der Faktorladungen sind hierbei sowohl nach Messzeitpunkt (Welle i) wie auch nach Klasse (1-3) aufgeschlüsselt, damit die Folgen der Invarianzformen auf die einzelnen Parameterwerte dargestellt werden können. Der Parameter y_{1i} stellt den Referenzindikator in jeder Konfiguration dar, welche die Skalierung der latenten Konstrukte festlegt, und wird somit nie verändert. In Tabelle 5 sind die negativen Intercept-Abweichungen für die längsschnittliche Non-Messinvarianz in den (Invar-

)Abstufungen 2 bis 3, sowie die positiv und negativ gerichteten Abweichungen für die gruppenspezifischen Unterschiede in Abstufung 4 und 5 abgebildet. In Abstufung 6 finden sich dann schließlich die kombinierten Werte für die längsschnittliche und gruppenspezifische Verzerrung. Analog hierzu, sind in Tabelle 4 die Unterschiede in den Faktorladungen für die Abstufungen 3, 5 und 6 abgebildet.

Auch aus Abbildung 6 können die Populationsparameter, erklärten Varianzanteile und Parameterbezeichnungen für die erste Form der Messinvarianz (INV) noch einmal überblicksartig entnommen werden.

8.3 Schätzung der Simulationsdaten

Für die Schätzung des Populationsmodells wird ein Modell spezifiziert, welches von invarianten Messmodellen über die Zeit, wie auch von einer invarianten Messung über die latenten Gruppen ausgeht. Dahinter steckt die Grundannahme, dass das Messinstrument für jede latente Gruppe zu jedem Zeitpunkt der Erhebung eine gleiche – oder mindestens annähernd gleiche – Bedeutung hat. Ohne diese Annahme wären Zeit- oder Gruppenvergleiche in der Forschungspraxis nicht verzerrungsfrei durchführbar. Im Schätzmodell wird ebenso angenommen, dass die Varianzen und Kovarianzen der Intercept- und Slopetermine über die Gruppen hinweg identisch sind, was so auch im Populationsmodell festgelegt wurde. Dies muss zwar in der Forschungspraxis nicht zwingend der Fall sein, wird aber oft aus Gründen von zu hoher Modellkomplexität und Konvergenzproblemen angenommen. Zur Skalierung der Varianz- und Mittelwertstruktur wird die «Marker Variable Approach» angewendet. Dabei wird die Faktorladung eines manifesten Indikators der latenten Konstrukte auf 1 (oder einen anderen konstanten Wert $\neq 0$) gesetzt. Analog hierzu wird der Intercept eines manifesten Indikators auf 0 fixiert. Dadurch nimmt das latente Konstrukt den Mittelwert bzw. die Skala der entsprechenden Indikatorvariablen an (vgl. Brown, 2006: 258; Wickrama, 2016: 88 ff.). Bei skalarer Faktorinvarianz von MI/SO-Wachstumsmodellen besteht die zusätzliche Besonderheit, dass Wertveränderungen der manifesten Variablen nicht auf das latente Konstrukt der Messgröße selbst, sondern auf die nächsthöhere Ebene der Wachstumsparameter verschoben werden (vgl. Urban, 2004: 52 f.).

Tabelle 4: Populationswerte der Faktorladungen für alle untersuchten Invarianzformen

Populationswerte: Faktorladungen (λ)															
Indikatoren	Wellen i Klasse 3					Wellen i Klasse 1					Wellen i Klasse 2				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
INV: Messinvarianz															
y_{1i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{2i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{3i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{4i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNI: laengssch. Nicht-Invarianz der Intercepts															
y_{1i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{2i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{3i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{4i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LNIF: laengssch. Nicht-Invarianz der Intercepts u. Faktorladungen															
y_{1i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{2i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{3i}	1.00	0.92	0.84	0.76	0.68	1.00	0.92	0.84	0.76	0.68	1.00	0.92	0.84	0.76	0.68
y_{4i}	1.00	0.92	0.84	0.76	0.68	1.00	0.92	0.84	0.76	0.68	1.00	0.92	0.84	0.76	0.68
GNI: gruppenspez. Nicht-Invarianz der Intercepts															
y_{1i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{2i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{3i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{4i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
GNIF: gruppenspez. Nicht-Invarianz der Intercepts u. Faktorladungen															
y_{1i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{2i}	1.00	1.00	1.00	1.00	1.00	1.00	0.70	0.70	0.70	0.70	1.00	1.15	1.15	1.15	1.15
y_{3i}	1.00	1.00	1.00	1.00	1.00	1.00	0.70	0.70	0.70	0.70	1.00	1.15	1.15	1.15	1.15
y_{4i}	1.00	1.00	1.00	1.00	1.00	1.00	0.70	0.70	0.70	0.70	1.00	1.15	1.15	1.15	1.15
LGNIF: kombinierte Nicht-Invarianz der Intercepts u. Faktorladungen															
y_{1i}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
y_{2i}	1.00	1.00	1.00	1.00	1.00	0.70	0.70	0.70	0.70	0.70	1.00	1.15	1.15	1.15	1.15
y_{3i}	1.00	1.00	1.00	1.00	1.00	0.70	0.62	0.54	0.46	0.38	1.15	1.07	0.99	0.91	0.83
y_{4i}	1.00	1.00	1.00	1.00	1.00	0.70	0.62	0.54	0.46	0.38	1.15	1.07	0.99	0.91	0.83

Info: y_{1i} stellt in jeder Klasse den Referenzindikator dar.

Tabelle 5: Populationswerte der Intercepts für alle untersuchten Invarianzformen

Populationswerte: Intercepts (τ)															
Indikatoren	Wellen i Klasse 3					Wellen i Klasse 1					Wellen i Klasse 2				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
INV: Messinvarianz															
y_{1i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{2i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{3i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{4i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LNI: laengssch. Nicht-Invarianz der Intercepts															
y_{1i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{2i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{3i}	0.00	-0.10	-0.20	-0.30	-0.40	0.00	-0.10	-0.20	-0.30	-0.40	0.00	-0.10	-0.20	-0.30	-0.40
y_{4i}	0.00	-0.10	-0.20	-0.30	-0.40	0.00	-0.10	-0.20	-0.30	-0.40	0.00	-0.10	-0.20	-0.30	-0.40
LNIF: laengssch. Nicht-Invarianz der Intercepts u. Faktorladungen															
y_{1i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{2i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{3i}	0.00	-0.10	-0.20	-0.30	-0.40	0.00	-0.10	-0.20	-0.30	-0.40	0.00	-0.10	-0.20	-0.30	-0.40
y_{4i}	0.00	-0.10	-0.20	-0.30	-0.40	0.00	-0.10	-0.20	-0.30	-0.40	0.00	-0.10	-0.20	-0.30	-0.40
GNI: gruppenspez. Nicht-Invarianz der Intercepts															
y_{1i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{2i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.25	0.25	0.25	0.25	0.25
y_{3i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.25	0.25	0.25	0.25	0.25
y_{4i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.25	0.25	0.25	0.25	0.25
GNIF: gruppenspez. Nicht-Invarianz der Intercepts u. Faktorladungen															
y_{1i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{2i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.25	0.25	0.25	0.25	0.25
y_{3i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.25	0.25	0.25	0.25	0.25
y_{4i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.25	0.25	0.25	0.25	0.25
LGNIF: kombinierte Nicht-Invarianz der Intercepts u. Faktorladungen															
y_{1i}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
y_{2i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.25	0.25	0.25	0.25	0.25
y_{3i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.60	-0.70	-0.80	-0.90	0.25	0.15	0.05	-0.05	-0.15
y_{4i}	0.00	0.00	0.00	0.00	0.00	-0.50	-0.60	-0.70	-0.80	-0.90	0.25	0.15	0.05	-0.05	-0.15

Info: y_{1i} stellt in jeder Klasse den Referenzindikator dar.

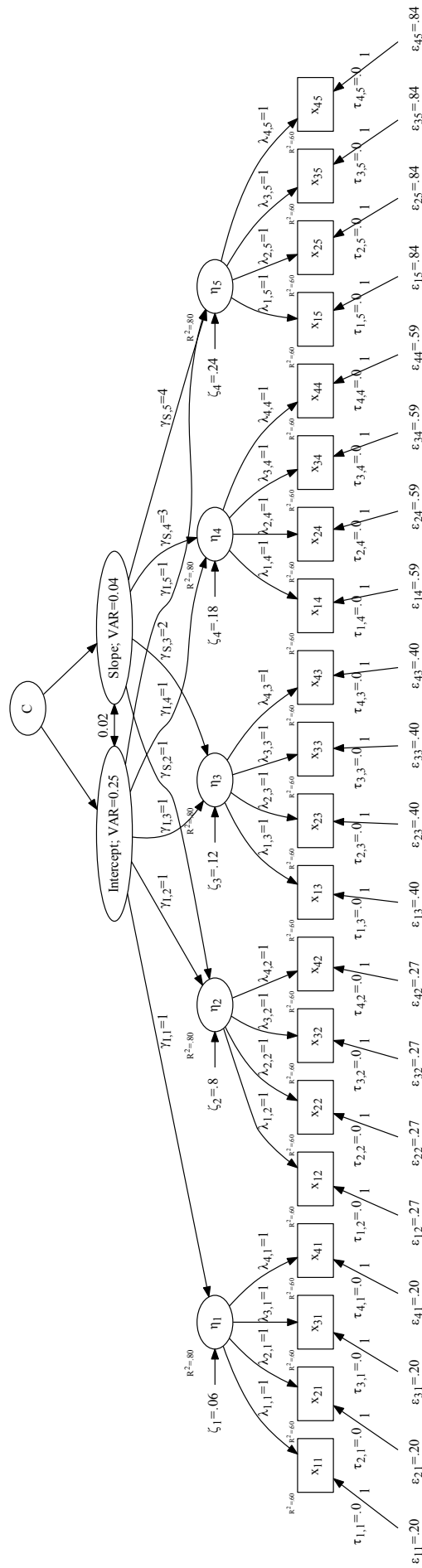


Abbildung 6: Populationswerte des (messtechnisch invarianten) MIGMM-Modells.

Diachrone Fehlerkorrelationen und die vollständige Mittelwertstruktur sind aus Übersichtlichkeitsgründen nicht abgebildet. Eigene Darstellung.

8.4 Zufallsziehung

Jede Ziehung des Monte-Carlo-Verfahrens erfolgt anhand eines computergesteuerten Zufallsprozesses. Die individuelle und idiosynkratische Stichprobenverteilung einer jeden Replikation enthält hiermit immer eine Zufallskomponente. Der damit verbundene Monte-Carlo-Fehler wird allerdings umso kleiner, je mehr Replikationen bzw. je mehr Stichproben aus dem DGP gezogen werden. Im Mittel gleichen sich dadurch die Eigenheiten jeder individuellen Ziehung aus. Nachvollziehbarkeit und Wiederholbarkeit der Simulation sind dadurch allerdings nur eingeschränkt möglich. Wenn der Zufallsprozess nicht wiederholt werden kann, dann ist nie ganz auszuschließen, dass es sich bei einer Abweichung im Rahmen einer Replikation der Simulationsstudie nicht doch um eine zufällige Abweichung handelt. Um die Nachvollziehbarkeit sicher zu stellen, besteht die Möglichkeit, den Zufallsprozess in seiner Gänze nachvollziehbar zu machen, indem der sogenannte Seed (auch: Seed Key) – der Startpunkt der Zufallsziehung – explizit angegeben wird. Da es sich bei computergenerierten Zufallszahlen immer (zumindest in aller Regel) um Pseudozufallszahlen handelt, kann die Reihenfolge der Zufallszahlen mit dem Initialisierungsseed nachvollzogen werden. Dafür wurde der Seed-Wert der Simulationsmodelle (willkürlich) auf 3454367 fixiert (siehe hierzu auch die Inputfiles des Onlineanhangs). Damit können die Ziehungen der Replikationen der Modelle innerhalb der jeweiligen Simulationskonfiguration konstant gehalten werden. Dies ist insbesondere deswegen relevant, weil innerhalb jeder Simulationskonfiguration eine Modellschätzung mit einer, zwei, drei und vier latenten Klassen durchgeführt werden muss, um überprüfen zu können, welche Klassenanzahl jeweils zur besten Lösung führt. Würde hierfür kein fester Seed verwendet werden, könnten die einzelnen Replikationen der Simulation nicht miteinander verglichen werden. Es wäre beispielsweise nicht klar, ob die Modellidentifikation auf die Zufallskomponente der Simulation oder auf inhärente Eigenschaften der MIGMM-Schätzmodelle zurückzuführen wäre.

8.5 Mixture-Model-Simulationen und Labelswitching

Bei Simulationsstudien von Mixture Modellen, wozu auch MIGMMs zählen, stellt sich das Problem, dass die Benennung der latenten Klassen als erste, zweite oder dritte (etc.) Klasse beliebig ist und von Statistikprogrammen wie Mplus auch nach diesem (beliebigen) Prinzip erfolgt (vgl. Tueller et al., 2011: 111 f.). Werden nun viele Replikationen

berechnet, kann es passieren, dass die Bezeichnung der Klassen zwischen den einzelnen Replikationen wechselt und somit die Parameterschätzungen nicht mehr sinnvoll über die Replikationen aggregiert werden können. In der Konsequenz führt dies dazu, dass Analysen zu Verzerrungen, welche auf Grundlage dieser aggregierten Parameterschätzungen berechnet werden, fehlerhaft sind. Analysen zur Klassenidentifikation (bzw. IC-Indizes und BLRT) sind davon allerdings nicht betroffen, da die Reihenfolge bzgl. der latenten Klassen hier keine Rolle spielt.

Um zu verhindern, dass bei Monte-Carlo-Replikationen ein Label-Wechsel zwischen verschiedenen Replikationen stattfindet, ist es üblich, benutzerdefinierte Startwerte für alle Parameter zu implementieren, um sicherzustellen, dass die Klassenbeschriftung über die Replikationen hinweg konsistent ist (vgl. Depaoli, 2013: 198). Aus diesem Grund wurden benutzerdefinierte Startwerte implementiert, die mit den Populationswerten übereinstimmen. Mögliche Probleme, die sich aus dem replikationsübergreifenden Label-Switching ergeben, konnten durch die Verwendung dieser benutzerdefinierten Startwerte vermieden werden. Die randomisierten Startwerte von Mplus, welche sicherstellen sollen, dass die Modellschätzungen keine lokalen Lösungen erzeugen, werden dabei als Abweichungen von diesen benutzerdefinierten Startwerten generiert (vgl. Muthén & Muthén, 2003). Zusätzlich können für die Modellschätzungen Identifizierbarkeits-Beschränkungen spezifiziert werden, welche die Möglichkeit für ein Label-Switching weiter verringern. In dieser Studie wurden Identifizierbarkeits-Beschränkungen innerhalb von Mplus auf der Grundlage der Mean-Werte der Intercept-Wachstumsfaktoren spezifiziert, so dass für die Meanwerte der latenten Klassen CX gilt $C1 > C2 > C3$.

8.6 Exkurs: EDV-technische Umsetzung der Simulationsberechnungen

Im Sinne der intersubjektiven Nachvollziehbarkeit (und möglichen Replikation) der Simulationsergebnisse wird in diesem kurzen Exkurs beschrieben, in welcher Form die Berechnungen im Detail umgesetzt wurden.

Als erster Schritt wurden die Mplus-Input-Files für die einzelnen Simulationskonfigurationen erstellt. Dabei wurde auf ein Template-File zurückgegriffen, welcher die einzelnen Modellvariationen der Simulation unter konditionalen Labels enthält. Mit Hilfe der Funktion «createModels» des R-Packets «MplusAutomation» wurden aus dem singulären

Template-File insgesamt 648 verschiedene Mplus-Input-Files erstellt. Darin sind die 162 verschiedenen Simulationskonfigurationen (Variablenkonfigurationen) mit den jeweils 4 (Klassen-)Modellschätzungen enthalten ($162 * 4$).

Im zweiten Schritt wurden diese 648 Input-Files zunächst auf die Verzeichnisstruktur des BwUniClusters hochgeladen. Um die Ziehung und die Berechnung der Replikation durchzuführen, wurden, im Anschluss daran, durch ein Shellskript-File automatisiert 648 einzelne «Job-Files» erstellt, welche alle nötigen Befehle und Parameter für die auf dem BwUniCluster eingesetzte Moab Cluster Suite (ein Cluster-Workload-Management-Package von Adaptive Computing, Inc) enthalten. Diese Job-Files wurden wiederum durch «Submit-Shellskripte» in 50er Stapeln an die Moab Cluster Suite zur Berechnung übergeben.³⁵

Das Ergebnis der Berechnung war jeweils ein (für jedes Input-File) allgemeines, über alle Replikationen aggregiertes Output-File von Mplus und ein zusätzliches «Results-File», welches die Parameterschätzungen und Kennwerte der einzelnen Replikationen beinhaltet. Diese Ergebnisdateien wurden wiederum von Clusterrechner heruntergeladen und als Rohdaten innerhalb der Statistiksoftwareumgebung R verarbeitet. Abgesehen von allgemeinen Informationen (wie z. B. zur Laufzeit) wurde dabei vorrangig auf die Informationen der einzelnen Replikationen in den «Result-Files» zurückgegriffen. Dieser Prozess wurde sowohl für die Hauptsimulation mit den IC-Indizes als auch für die Zusatzsimulation mit der BLRT durchlaufen.³⁶

³⁵Die Nutzerrechte des BwUniCluster lassen es nicht zu, dass die 648 Einzelberechnungen gleichzeitig durchgeführt werden. Daher musste dies in 50er Stapeln umgesetzt werden.

³⁶Für die Simulationen zum BLRT mussten pro Konfiguration und Replikation, allerdings, lediglich drei Modelle geschätzt werden, da die 2-Klassenmodelle den Test gegenüber den 1-Klassenmodellen bereits enthalten und ein 0-Klassenmodell logisch nicht möglich ist.

9 Ergebnisse zur Klassenidentifikation

9.1 Haupteffekte bzgl. AIC, BIC, aBIC

Zur Analyse der Klassenidentifikation werden die Einflüsse der Simulationsvariablen zunächst als einzelne Haupteffekte dargestellt. Im Sinne eines unifaktoriellen Designs wird zunächst nur die betrachtete Simulationsgröße variiert. Um dabei einen möglichst unverfälschten Haupteffekt zu betrachten, werden die Ausprägungen der jeweils nicht betrachteten Variablen konstant gehalten. Zusätzlich werden diese nicht betrachteten Variablen auf einen Wert fixiert, welcher einen möglichst geringen Einfluss auf die Klassenidentifikation haben sollte, damit der Einfluss der unabhängigen Größen möglichst isoliert (und von möglichen Interaktionen unverfälscht) betrachtet werden kann.

Dafür wird für die Invarianzstufe die Ausprägung «1» ausgewählt, welche von einer vollständigen skalaren Invarianz der Messung bzgl. der Faktorladungen und Intercepts der manifesten Indikatoren ausgeht. Die Messinvarianz ist in dieser Simulationsausprägung sowohl über die Zeit als auch über die Gruppen hinweg gegeben. Für die Fallzahl (nobs) der einzelnen Replikationen wird die Ausprägung auf «3» fixiert, was einer Fallzahl von 1300 simulierten Beobachtungen entspricht. Die bisherigen Ergebnisse zu GM-Modellen zeigen eindeutig, dass eine hohe Fallzahl die korrekte Klassenidentifikation, in aller Regel, begünstigt. Für die Klassendistanz (sep) wird die Ausprägung «3» angesetzt, was einer Klassendistanz von 4.0 MD entspricht und allgemein als eine sehr hohe Klassendistanz zu werten ist. Nach bisherigen Simulationsstudien stellt eine hohe Klassendistanz eine zentrale Bedingung für die richtige Klassenidentifikation dar. Für das Mengenverhältnis zwischen den Klassen wird die Ausprägung «1» gewählt, was einem gleichen prozentualen Anteil bzgl. der Fallzahlen innerhalb der Klassen (33 %, 33 %, 33 %) entspricht.

Tabelle 6: Prozentuale Klassenidentifikation nach Invarianz-Stufen

invar	1.aic	2.aic	3.aic	4.aic	1.bic	2.bic	3.bic	4.bic	1.abic	2.abic	3.abic	4.abic
1.00	0.00	0.00	64.33	35.67	0.00	0.00	100.00	0.00	0.00	0.00	95.00	5.00
2.00	0.00	0.00	68.67	31.33	0.00	0.00	100.00	0.00	0.00	0.00	97.00	3.00
3.00	0.00	0.00	67.00	33.00	0.00	0.00	100.00	0.00	0.00	0.00	96.00	4.00
4.00	0.00	0.00	31.00	69.00	0.00	0.00	80.00	20.00	0.00	0.00	54.00	46.00
5.00	0.00	0.00	8.33	91.67	0.00	32.00	41.67	26.33	0.00	2.33	26.67	71.00
6.00	0.00	7.00	28.67	64.33	0.00	90.67	8.00	1.33	0.00	50.33	30.33	19.33

Info: Klassendistanz=4 MD, Mixing=33.33.33, Nobs=1300.

Fallzahl pro Konfiguration jeweils 300 (Replikationen)

Invarianzstufen (invar). In Tabelle 6 ist zu erkennen, dass die verschiedenen Abstufungen der (Non-)Messinvarianz einen unterschiedlichen Effekt auf die Klassenidentifikation haben. Zunächst fällt auf, dass unabhängig vom zugrundeliegenden IC-Index der prozentuale Anteil der als 3-Klassenmodell richtig identifizierten Replikationen für die Invarianzabstufungen 1 bis 3 sehr konstant bleibt. Dies zeigt, dass – unter ansonsten weitestgehend optimalen (Simulations-)Bedingungen – die längsschnittliche Non-Invarianz weder hinsichtlich der Intercepts noch hinsichtlich der Faktorladungen einen negativen Einfluss auf die Klassenidentifikation (nach den betrachteten IC-Indizes) aufweist.

Unter den optimalen Randbedingungen der hier spezifizierten Simulationen zeigt sich zugleich, dass BIC und aBIC eine deutlich bessere Klassenidentifikation aufweisen als der AIC, welcher in der Tendenz häufig mehr Klassen identifiziert, was mit bisherigen Ergebnissen im Einklang ist (vgl. Nylund et al., 2007: 537). Während die Rate der korrekten Klassenidentifikation für BIC und aBIC auch unter Verletzung der längsschnittlichen Messinvarianz im Bereich von 95 bis 100 % bleibt, bewegt sich der AIC im Bereich von 64 bis 69 % und schneidet damit eindeutig schlechter ab, was allerdings, wie erwähnt, mit bisherigen Ergebnissen im Rahmen von Simulationsstudien zu Growth Mixture Modellen übereinstimmt.

Für die 4. und 5. Messinvarianzkonfiguration, welche Non-Invarianz zwischen den latenten Gruppen hinsichtlich Intercepts und Faktorladungen repräsentieren, kann nicht mehr von Stabilität oder Konstanz gesprochen werden. Die Trefferquoten aller untersuchten IC-Indizes werden klar beeinflusst. Fehlende Messinvarianz hinsichtlich der Intercepts wirkt sich für AIC, BIC und aBIC jeweils gleichgerichtet aus, indem deutlich häufiger die 4-Klassenlösung als richtig identifiziert wird. Während AIC und aBIC mit einer Verschiebung um ca. 40 Prozentpunkte reagieren, zeigt sich der BIC mit einer Anteilsverschiebung um 20 Prozentpunkte etwas weniger sensitiv hinsichtlich der fehlenden Intercept-Messinvarianz. Dieses Ergebnis kann damit plausibilisiert werden, dass die Nichtberücksichtigung von fehlender Intercept-Messinvarianz bei der Modellschätzung eine zusätzliche Komponente von Heterogenität zwischen den Klassen einführt, welche über allgemeine Klassendistanz hinaus geht (vgl. Kim & Wang, 2017: 8). Insofern ist es plausibel, dass diese zusätzliche Heterogenität mit einer häufigeren Identifikation von zu vielen Klassen einhergeht.

Dieser Effekt wird in der 5. Konfiguration für den AIC und den aBIC noch weiter verstärkt. Die zusätzlich fehlende Messinvarianz der Faktorladungen führt hier zu einer

noch stärkeren Überidentifikation von latenten Klassen. Die Klassifikation nach dem BIC fällt etwas aus dem Rahmen, da sich hier sowohl eine Über- wie auch eine Unteridentifikation von latenten Klassen feststellen lässt. Auch hier zeigt sich allerdings der BIC noch am robustesten hinsichtlich der korrekten Klassenidentifikation, da er noch in knapp 42 % der Replikationen die korrekte Klassenanzahl anzeigen kann, wohingegen AIC und aBIC deutlich schlechter abschneiden.

In der 6. Simulationskonfiguration der Messinvarianz werden längsschnittliche und gruppenspezifische Non-Invarianzen hinsichtlich der Intercepts und der Faktorladungen kombiniert. Die gruppenspezifische Non-Invarianz wurde in einer Gruppe als eine positive und in der anderen Gruppe als eine negative Abweichung spezifiziert. Da die längsschnittliche Abweichung konstant als negative spezifiziert wurde, nähert sich die positive Abweichungsgruppe im Schnitt an die Referenzgruppe an, während sich die negative Abweichungsgruppe weiter von der Spezifikation der Referenzgruppe entfernt (siehe hierzu auch Tabelle 4 und 5).

Wie vermutet, zeigt sich eine Interaktion zwischen fehlender längsschnittlicher und gruppenspezifischer Messinvarianz. Während die gruppenspezifische Non-Invarianz in Stufe 5 in der Tendenz zur Identifikation von zu vielen Klassen führte, dreht sich dieser Trend bei der Kombination beider Formen von fehlender Messinvarianz um. BIC und aBIC favorisieren die 2-Klassenlösung deutlich häufiger. Für den AIC nimmt der Anteil der richtigen Klassen zwar zu, was aber unter Berücksichtigung der Tendenz zur Überidentifikation des AIC einem gleichgearteten Effekt wie beim BIC und aBIC entspricht. Durch die Annäherung der einen latenten Klasse an die Referenzgruppe (und der gleichzeitigen weiteren Entfernung der anderen Klasse) werden im Durchschnitt eher Lösungen favorisiert, welche weniger Klassen aufweisen.

Anhand der Klassenidentifikation entlang der sechs Abstufungen der Simulation zur Messinvarianz lassen sich drei wesentliche Ergebnisse ableiten.

- Alle IC-Indizes sind äußerst robust gegenüber längsschnittlicher Non-Invarianz. Dies gilt sowohl für die Faktorladungen als auch für die Intercepts der manifesten Indikatoren.
- Gruppenspezifische Non-Invarianz wirkt sich deutlich negativ auf die Fähigkeit der IC-Indizes zur korrekten Klassenidentifikation aus. Dies gilt bereits bei Verzerrung

der Intercepts und wird zusätzlich durch fehlende Invarianz hinsichtlich der Faktorladungen verstärkt.

- In Kombination mit gruppenspezifischen Unterschieden wirkt sich auch fehlende längsschnittliche Invarianz der Messung auf die Klassenidentifikation nach den IC-Indizes aus. Je nach Stoßrichtung der Vorzeichen werden Gruppenunterschiede verstärkt oder nivelliert, was die Klassenidentifikation verfälscht.

Tabelle 7: Prozentuale Klassenidentifikation nach Klassendistanz-Stufen

sep	1.aic	2.aic	3.aic	4.aic	1.bic	2.bic	3.bic	4.bic	1.abic	2.abic	3.abic	4.abic
1.00	0.00	75.67	20.67	3.67	0.00	100.00	0.00	0.00	0.00	99.00	1.00	0.00
2.00	0.00	60.33	30.00	9.67	0.00	100.00	0.00	0.00	0.00	96.00	4.00	0.00
3.00	0.00	0.00	64.33	35.67	0.00	0.00	100.00	0.00	0.00	0.00	95.00	5.00

Info: Invar=1, Mixing=33.33.33, Nobs=1300.

Fallzahl pro Konfiguration jeweils 300 (Replikationen)

Klassendistanz (sep). Die Ergebnisse zur Klassendistanz sind sehr eindeutig (siehe Tabelle 7). Für die Klassenunterschiede von 0.8 MD (sep=1) und 1.5 MD (sep=2) wird durchgehend eine Klasse zu wenig identifiziert. Während dies für den BIC in 100 % der Replikationen der Fall ist, schneidet der aBIC mit 96 und 99 % der Replikationen nur marginal besser ab. Der AIC schneidet, aufgrund seiner allgemeinen Bevorzugung von Lösungen mit mehr Klassen, hier etwas besser ab. Trotz dessen wird in der Mehrzahl der Replikationen nicht die korrekte Anzahl an Klassen identifiziert.

Für die Simulationsausprägung mit einer Klassendistanz von 4.0 MD (sep=3) dreht sich das Bild völlig. BIC liefert über die Replikationen hinweg ein perfektes Ergebnis ab, während der aBiC in 95 % der Fälle die korrekte 3-Klassenlösung findet. Der AIC schneidet unter einer großen Klassendistanz wiederum vergleichsweise schlecht ab und findet nur in knapp 64 % der Replikation die richtige Lösung, wobei die nicht richtig identifizierten Replikationen zugunsten von zu vielen Klassen ausfallen. Hier zeigt sich, wie zentral die Klassendistanz für die korrekte Identifikation von latenten Klassen ist. Obwohl sich die Klassendistanzen von 0.8 und 1.5 MD im Rahmen der üblichen Werte für Simulationsstudien zu GMMs bewegen (vgl. z. B.: Depaoli, 2013; Liu & Hancock, 2014; Lubke & Neale, 2006; Peugh & Fan, 2012), zeigt sich, dass für die Klassenidentifikation in Rahmen von MIGM-Modellen höhere Klassendistanzen anzuraten sind, um verlässliche Ergebnissen zu bekommen.³⁷

³⁷Natürlich gilt dieser Befund strenggenommen nur für MIGMMs, welche in ähnlicher Form (z. B. mit diachronen Residuenkorrelationen) spezifiziert wurden, wie das für die Simulation angesetzte Simulati-

Tabelle 8: Prozentuale Klassenidentifikation nach Nobs-Stufen

nobs	1.aic	2.aic	3.aic	4.aic	1.bic	2.bic	3.bic	4.bic	1.abic	2.abic	3.abic	4.abic
1.00	0.00	0.00	61.33	38.67	0.00	1.00	98.33	0.67	0.00	0.00	73.33	26.67
2.00	0.00	0.00	60.00	40.00	0.00	0.00	99.67	0.33	0.00	0.00	89.33	10.67
3.00	0.00	0.00	64.33	35.67	0.00	0.00	100.00	0.00	0.00	0.00	95.00	5.00

Info: Invar=1, Mixing=33.33.33, Klassendistanz=4 MD.
Fallzahl pro Konfiguration jeweils 300 (Replikationen)

Fallzahl (nobs). Aus Tabelle 8 können die Ergebnisse für die Variation der Fallzahl zwischen 300 (nobs=1), 700 (nobs=2) und 1300 (nobs=3) Beobachtungsfällen entnommen werden. Der BIC zeigt für jede Fallzahlkombination äußerst robuste Klassifikationsanteile an. Zwar steigt auch für den BIC die Rate der richtigen Identifikationen mit der Fallzahl, aber die Unterschiede sind hier so gering, dass sie kaum noch interpretationsfähig sind, wenn man berücksichtigt, dass ein gewisser Grad an Variabilität bereits durch die Zufallsziehung der Replikationen gegeben ist. Für den aBIC gestalten sich die Unterschiede schon deutlicher. Für 300 Beobachtungsfälle werden nur knapp 74 % der Replikationen richtig identifiziert, was bei 1300 Beobachtungsfällen bereits auf 95 % ansteigt. Der AIC zeigt sich ebenfalls hinsichtlich der Fallzahl sensibel, auch wenn der Effekt deutlich geringer als bei aBIC ausfällt und auf einer niedrigeren Ausgangsbasis (zwischen 60 und 64 %) verläuft.

Es kann festgehalten werden, dass (unter ansonsten günstigen Randbedingungen, wie Mesinvarianz, gleichen Gruppenanteilen und guten Klassenseparationen) die Fallzahl einen positiven Einfluss auf die Identifikationsraten des aBIC und, in geringerem Maße, auf den AIC ausübt. Der BIC scheint hingegen schon ab der niedrigsten getesteten Fallzahl von 300 gute Ergebnisse zu liefern. Die hier identifizierte gesteigerte Sensitivität des aBIC gegenüber der Fallzahl zeigte sich auch in anderen Studien (konkret unter 500 Fällen) (vgl. Henson et al., 2007: 221 ff.).

Tabelle 9: Prozentuale Klassenidentifikation nach Mixing-Stufen

mixing	1.aic	2.aic	3.aic	4.aic	1.bic	2.bic	3.bic	4.bic	1.abic	2.abic	3.abic	4.abic
1.00	0.00	0.00	64.33	35.67	0.00	0.00	100.00	0.00	0.00	0.00	95.00	5.00
2.00	0.00	0.00	62.67	37.33	0.00	0.00	100.00	0.00	0.00	0.00	95.67	4.33
3.00	0.00	0.00	68.00	32.00	0.00	3.33	96.67	0.00	0.00	0.00	94.00	6.00

Info: Invar=1, Nobs=1300, Klassendistanz=4 MD.
Fallzahl pro Konfiguration jeweils 300 (Replikationen)

Mixing Proportions (mixing). Der Effekt der Anteilsverhältnisse zwischen den Klassen fällt insgesamt sehr schwach aus. Die gleichverteilte Klassenzusammensetzung (mixing=1) von jeweils 33 % der Beobachtungen pro Klasse und die moderat schiefverteilte Klassenzusammensetzung (mixing=2) mit 17, 33 und 40 % der Beobachtungsfälle weisen hinsichtlich der Klassenidentifikation von AIC, BIC und aBIC keine interpretationsfähigen Unterschiede auf (siehe Tabelle 9). Erst die dritte Simulationskonfiguration (mixing=3) mit einer kleinen Klasse von 5 % (5 %, 45 %, 50 %) zeigt geringfügige Unterschiede hinsichtlich der Trefferquote von AIC und BIC. Für den AIC steigt der Anteil der korrekt zugeordneten Replikationen auf Kosten des Anteils an 4-Klassenlösungen. Hierbei liegt die Interpretation nahe, dass die allgemeine Tendenz des AIC zur Identifikation von zu vielen Klassen bei einer kleinen Klassengröße kompensiert wird. Beim insgesamt hinsichtlich der Klassenidentifikation als deutlich konservativer einzuschätzenden BIC (vgl. Usami et al., 2017: 594) zeigt sich der gegenteilige Zusammenhang. Klassenanteilsverhältnisse mit einer besonders kleinen Klasse führen hier zu einer leichten Tendenz zur Identifikation von 2-Klassenlösungen, wobei auch diese Tendenz sehr schwach ausgeprägt ist. Abweichungen der Trefferquote des aBIC erscheinen nicht interpretationswürdig.

Die fehlenden Unterschiede zwischen gleich und moderat ungleich verteilten Konfigurationen sind nicht überraschend und wurden so auch schon bei der Begründung des Simulationsdesigns vermutet. Bisherige Simulationsstudien haben nahe gelegt, dass moderat ungleichgewichtete Klassen keinen wesentlichen Effekt auf die Modellperformanz haben (vgl. Liu & Hancock, 2014: 190). Vermeintlich überraschend erscheint allerdings der geringe Einfluss in der dritten Kategorie mit einer äußerst kleinen Klasse. Hier war zu erwarten, dass sich die Chance für die Identifikation der korrekten Klassenanzahl deutlich verschlechtert (vgl. Depaoli, 2013: 190). Hierbei muss allerdings beachtet werden, dass der geringe Effekt vor dem Hintergrund der anderen fixierten Simulationsparameter betrachtet werden muss. Insbesondere die hohe Fallzahl von 1300 ist hierbei entscheidend. Auch bei einem prozentual geringen Klassenanteil von 5 % ist die kleine Klasse noch mit durchschnittlich (im Rahmen der Schwankungen der Replikationen) 65 Beobachtungen vertreten, was die erfolgreiche Klassenidentifikation begünstigt. Wie auch schon in der Diskussion der Simulationsparameter angemerkt, wird sich der Effekt der Konfiguration mit einer kleinen Klasse erst in Kombination mit anderen Simulationsgrößen (wie der absoluten Fallzahl) voll entfalten (vgl. hierzu auch Abschnitt 9.2).

9.2 Interaktionseffekte bzgl. AIC, BIC, aBIC

Neben den isolierten Haupteffekten der einzelnen Simulationsgrößen wurden auch Annahmen hinsichtlich der Kombination der Simulationsparameter aufgestellt. Diese Interaktionseffekte sollen im Weiteren untersucht werden. Dafür werden wiederum, wie zuvor für die Haupteffekte, die jeweils nicht betrachteten Simulationsgrößen konstant gehalten, um eine möglichst verzerrungsfreie Interpretation zu ermöglichen. Zur Darstellung der Ergebnisse wurde auf sogenannte Mosaikplots zurückgegriffen, welche die mehrfachen Kontingenztabellen leichter interpretierbar wiedergeben können. Die Flächen der rechteckigen Felder, die für eine Merkmalskombination stehen, sind in ihrer Größe proportional zur Anzahl der Beobachtungen/Replikationen, die diese Merkmalskombination aufweisen. Die schwarzen Rechtecke (sofern überhaupt sichtbar/vorhanden) stehen für die Anzahl der Replikationen, welche als 1-Klassenmodell eingestuft wurden. Die roten Rechtecke stehen für die 2-Klassenlösungen. Grün steht hingegen für die korrekte Identifikation als 3-Klassenmodell, wie es im Populationsmodell vorgesehen ist. Die Größe der blauen Rechtecke gibt hingegen die Anzahl der 4-Klassenlösungen wieder. Die proportionalen Anteile der Rechtecke können miteinander verglichen und in Beziehung gesetzt werden, da die zugrundeliegende Fallzahl immer 300 Replikationen entspricht und die übrigen Simulationsparameter auf analytisch günstigen Ausprägungen konstant gehalten werden. Die genaue prozentuale (numerische) Aufschlüsselung der Klassenidentifikation kann zum Vergleich auch aus dem Anhang entnommen werden (siehe z. B. Tabelle 41 und 42).

Fallzahl und Mixing Proportions. Es wurde angenommen, dass analog zu den bisherigen Ergebnissen zu GMMs die Mixing Proportions, für sich genommen, wenig Einfluss auf die Klassenidentifikation von MIGMMs ausüben. Lediglich in Datensituationen, in welchen sich ungünstige Bedingungen addieren, wurde ein Effekt vermutet. Insbesondere die Kombination aus geringer Fallzahl und kleinen prozentualen Klassengrößen sollte einen Effekt aufweisen.

Wie Grafik 7 entnommen werden kann, zeigt sich der vermutete Effekt so auch im Ergebnis der Simulation. Der Anteil der Replikationen, welche falsch als 2-Klassenmodelle identifiziert wurden (rote Rechtecke in Grafik) zeigt sich fast nur in der Mixing-Proportions-Ausprägung mit einer sehr kleinen Klasse und zudem nur, wenn die Fallzahl der Simulation bei 300 oder 700 Beobachtungen liegt.

Daraus lässt sich der Schluss ziehen, dass Mixing Proportions nur bei einer sehr kleinen

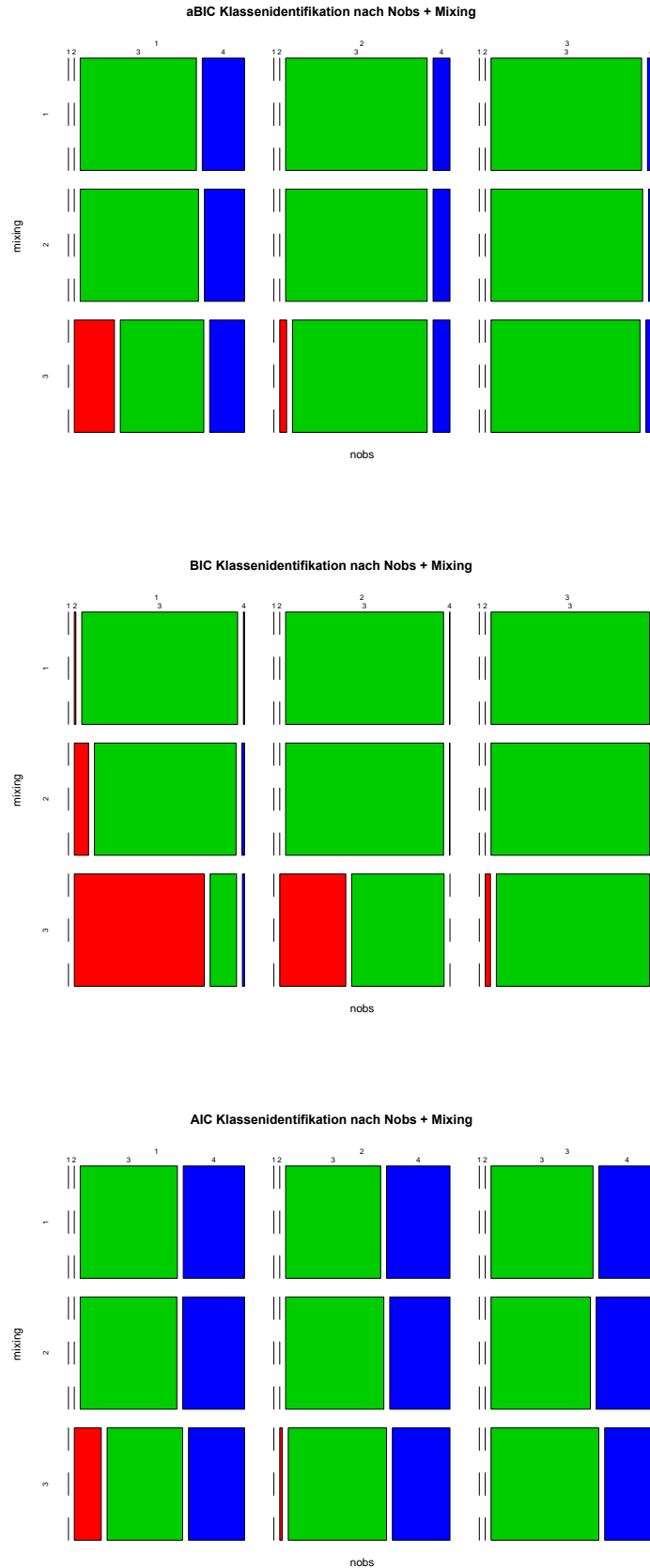


Abbildung 7: Interaktionseffekt zwischen Fallzahl und Klassenanteilen (Nobs*Mixing) bzgl. Klassenidentifikation nach AIC, BIC und aBIC.

Replikationsanzahl pro Konfiguration jeweils 300.

Legende: Schwarz=1-Klassen, Rot=2-Klassen, Grün= 3-Klassen; Blau=4-Klassen

Klasse (5 %) für die Klassenidentifikation von MIGMMs relevant sind und dieses Problem zudem durch eine hohe Fallzahl (hier 1300) so gut wie vollständig behoben werden kann. Auch wenn diese Beziehung sowohl für AIC, BIC und aBIC gilt, scheint die hohe Fallzahl, bei einem ungünstigen Klassenverhältnis, besonders für den BIC wichtig zu sein. Wie bei der Interpretation des Haupteffekts der Mixing Proportions bereits vermutet, zeigt sich der Effekt von prozentualen kleinen Klassen nur bei einer ebenfalls geringen Fallzahl. Für die gleichgewichteten und moderat ungleichgewichteten Klassenanteile lässt sich kaum ein Interaktionseffekt mit der Fallzahl ausmachen.³⁸

Messinvarianz und Klassendistanz. Die Ergebnisse der kombinierten Betrachtung von Messinvarianzausprägungen und den Abstufungen der Klassenseparation spiegeln zum Großteil die Ergebnisse der bereits betrachteten Haupteffekte wieder (vgl. Abbildung 8). Der starke Einfluss der Klassendistanz zeigt sich in der Überrepräsentation der 2-Klassenlösung für die Distanzen 0.8 MD (1) und 1.5 MD (2). Diese Dominanz kann für BIC und aBIC nur dann aufgebrochen werden, wenn durch die gruppenspezifischen Nicht-Invarianzen eine zusätzliche Form von Klassen-Heterogenität eingeführt wird. Dies trifft für den aBIC in einem höheren Maße zu, wie für den BIC. Betrachtet man die Ergebnisse von AIC im Vergleich, kann man zu der Schlussfolgerung verleitet werden, dass der AIC bei einer geringeren Klassendistanz, in Kombination mit partieller längsschnittlicher oder gruppenspezifischer Non-Invarianz, besser geeignet ist als die anderen IC-Maße. Allerdings zeigt der AIC bei geringer Klassendistanz in den stärker ausgeprägten Non-Invarianzkonfigurationen eine massive Überidentifikation hin zur 4-Klassenlösung. In Anbetracht dessen, kann in dieser Gegenüberstellung kein klarer Gewinner zwischen den IC-Maßen ausgemacht werden. Das Zusammenspiel zwischen der Klassendistanz und den Invarianzstufen läuft, alles in allem, darauf hinaus, dass sich in der invarianten und in den längsschnittlichen Konfigurationen zunächst kein Moderationseffekt zeigt und erst bei der Betrachtung der Konfigurationen mit Gruppenunterschieden eine Tendenz zur Identifikation von mehr Klassen sichtbar wird. Wie man am aBIC und am AIC erkennen kann, führt diese Tendenz jedoch nicht zwangsläufig zu einer besseren Klassenidentifikation, sondern kann auch zur Überextraktion führen. Für die Kombination von längsschnittlicher und gruppenspezifischer Non-Invarianz lässt sich zudem keine klare Tendenz des Interaktionseffektes erkennen.

³⁸Nur der BIC reagiert in der Kategorieausprägung der moderat ungleichgewichteten Klassenanteile auf eine Fallzahl von 300 mit einer geringfügig schlechteren Klassenidentifikation.

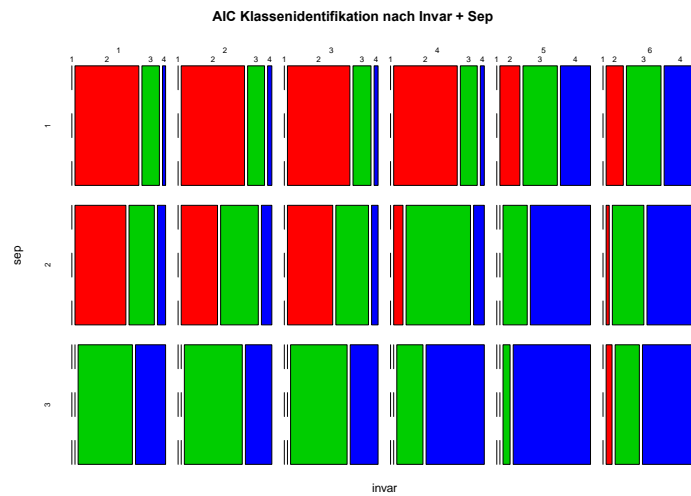
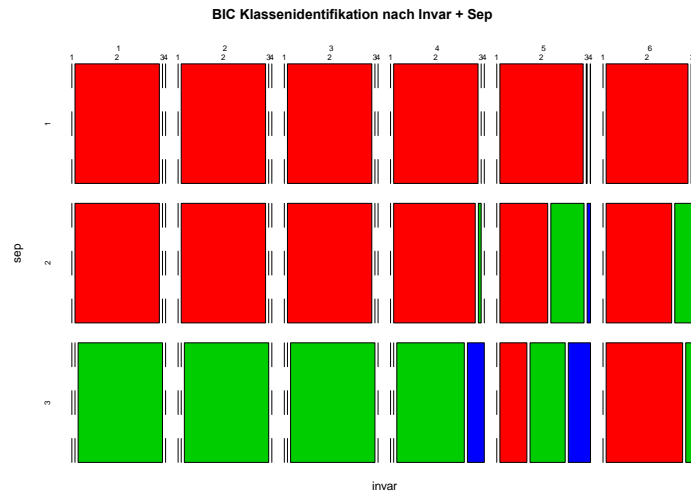
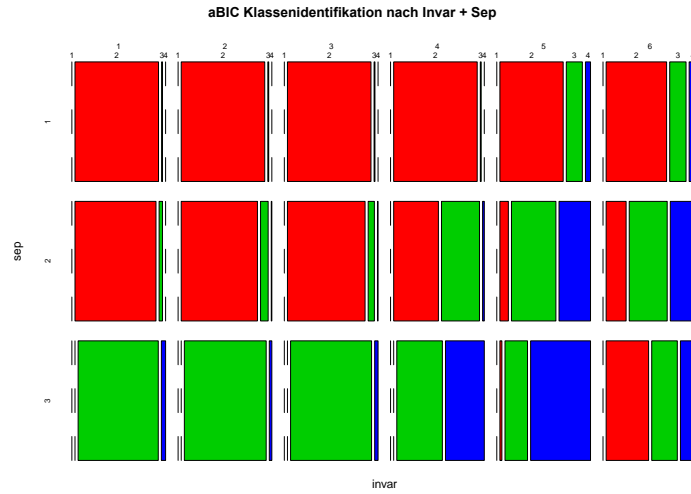


Abbildung 8: Interaktionseffekt zwischen Invarianzabstufungen und Klassendistanz (Invar*Sep) bzgl. Klassenidentifikation nach AIC, BIC und aBIC.

Replikationsanzahl pro Konfiguration jeweils 300.

Legende: Schwarz=1-Klasse, Rot=2-Klassen, Grün= 3-Klassen; Blau=4-Klassen

Messinvarianz und Fallzahl. Wie bereits bei der Betrachtung des Haupteffekts der Fallzahl zeigen sich bei der kombinierten Analyse nach Invarianzform und Fallzahl Unterschiede zwischen den IC-Indizes (vgl. Abbildung 9). Der anfangs angenommene verstärkende negative Einfluss von kleinen Fallzahlen auf Invarianzverletzungen lässt sich in dieser allgemeinen Form nicht halten.

Der BIC zeigt in den längsschnittlichen Invarianzenformen keine nennenswerte Sensitivität hinsichtlich der Fallzahl. Für alle Invarianz- und Fallzahlkombinationen ist die korrekte Klassenidentifikation sehr hoch. Erst ab den gruppenspezifischen Messunterschieden lässt sich ein Effekt ausmachen. Für Gruppenunterschiede in den Interceptwerten (Invar=4) zeigt sich mit steigender Fallzahl eine leichte Tendenz zur Überidentifikation von latenten Klassen zugunsten der 4-Klassenlösung. Für die Konfiguration mit Gruppenunterschieden in den Intercepts und Faktorladungen bleibt diese Tendenz zur Überidentifikation erhalten, wirkt sich aber in der Summe positiv auf die korrekte Klassenidentifikation aus, da häufiger die richtige 3-Klassenlösung bevorzugt wird. Das gleiche Muster zeigt sich auch in der kombinierten Konfiguration mit längsschnittlicher und gruppenspezifischer Nicht-Invarianz, wenn auch weniger stark ausgeprägt und in absoluten Zahlen auf einem niedrigen Niveau. Insgesamt lässt sich festhalten, dass für den BIC in den längsschnittlichen Invarianzenformen kein interpretationswürdiger Interaktionseffekt mit der Fallzahl auftritt. Die Wirkung von Gruppenunterschieden und kombinierten Unterschieden in den Messmodellen kann allerdings durch eine hohe Fallzahl abgemildert werden, wodurch häufiger richtige Klassenlösungen identifiziert werden können. Dieser Interaktionseffekt ist aber bei weitem nicht so stark, dass der Haupteffekt von fehlender Invarianz in Abstufung 4 bis 6 aufgehoben werden kann.

Für den aBIC zeigt sich ein einheitliches Muster in der invarianten und den beiden längsschnittlich abweichenden Konfigurationen (1 bis 3). Je höher die Fallzahl, desto eher wird die korrekte Klassenanzahl statt der falschen 4-Klassenlösung festgestellt. Der Umfang dieses Effekts unterscheidet sich zwischen den ersten drei Invarianzstufen nicht, weshalb noch nicht von einem Interaktionseffekt zu sprechen ist. Hier schlägt lediglich der Haupteffekt der Fallzahl auf die Klassenidentifikation des aBIC durch. Die 4. Invarianzkonfiguration stellt für diese Beziehungen jedoch den Wendepunkt dar. Hier spielt die Fallzahl zunächst keine Rolle für die Klassenidentifikation. In der 5. und 6. Stufe mit den Messunterschieden in den Intercepts und Faktorladungen erhöht sich hingegen mit steigender Fallzahl der Anteil der 4-Klassenlösungen. Das gleiche Fazit lässt sich auch auf

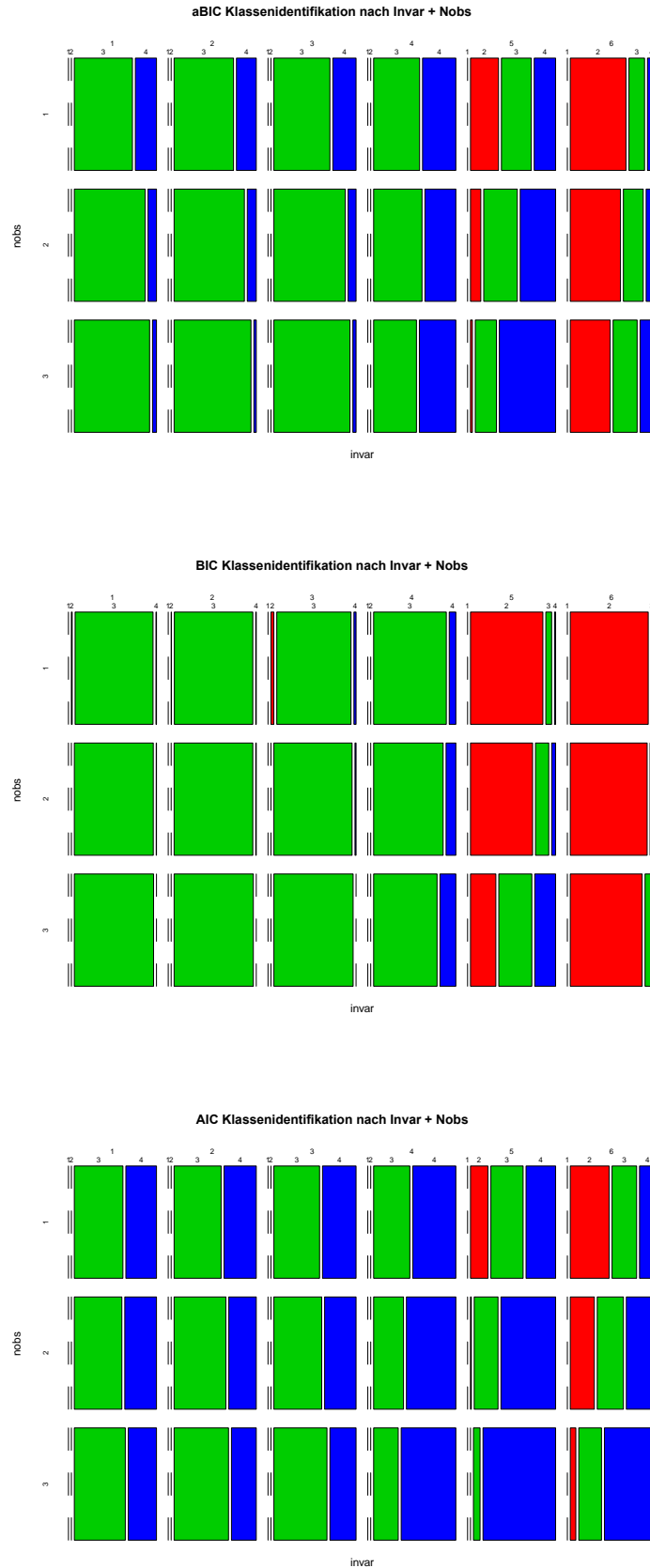


Abbildung 9: Interaktionseffekt zwischen Invarianzabstufungen und Fallzahl (Invar*Nobs) bzgl. Klassenidentifikation nach AIC, BIC und aBIC.

Replikationsanzahl pro Konfiguration jeweils 300.

Legende: Schwarz=1-Klasse, Rot=2-Klassen, Grün= 3-Klassen; Blau=4-Klassen

den AIC ausweiten. Der einzige Unterschied im Vergleich zum Zusammenhangsmuster für den aBIC besteht darin, dass der AIC in der invarianten und den längsschnittlichen Konfiguration nicht von der Fallzahl (als Haupteffekt) beeinflusst wird.

Zusammenfassend lässt sich für die drei IC-Indizes kein Interaktionseffekt der Fallzahl in den längsschnittlichen Non-Invarianzausprägungen ausmachen. Erst in den Konfigurationen mit Gruppenunterschieden lässt sich eine Tendenz zur Identifikation von mehr latenten Klassen ausmachen. Dies kann sich in der Summe leicht positiv auf die korrekte Klassenidentifikation auswirken (siehe BIC), aber auch zur vermehrten Überidentifikation von unnötigen Klassen führen (siehe AIC und aBIC).

Messinvarianz und Mixing. Für das Zusammenspiel von Klassengröße und Verletzungen der Messinvarianz wurde angenommen, dass sich Interaktionen vor allem für die Konfiguration mit einer sehr kleinen Klasse einstellen sollten. Nach ersten Plausibilitätsüberlegungen sollte sich durch die Kombination von kleiner Klassengröße und Verzerrungen in den Messungen die Chance dafür erhöhen, dass die kleine Klasse einer anderen Entwicklungsklasse zugeordnet wird. Auch der entgegengesetzte Zusammenhang ist denkbar, bei dem die kleine Klasse bei fehlerbehafteter Messung in zwei distinkte Gruppen eingruppiert wird oder zumindest die Abgrenzung der anderen latenten Wachstumsverlaufsgruppen erschwert. Je nach empirischer Verteilung (bzw. je nach simulierter Zufallsziehung) erscheinen beide Stoßrichtungen als plausibel.

Für die ersten beiden Invarianzstufen (Invar 1 u. 2) zeigt sich für keinen der IC-Indizes ein bedeutender Unterschied in der Klassenidentifikation in Abhängigkeit von den prozentualen Klassenverhältnissen (vgl. Abbildung 10). Erst in der dritten Invarianzstufe mit längsschnittlichen Unterschieden in den Intercepts und den Faktorladungen lässt sich eine Interaktion für alle drei IC-Indizes identifizieren. Für den aBIC und den AIC führt die Kombination aus längsschnittlicher Non-Invarianz und kleiner Klassengröße zu einer deutlich stärkeren Überidentifikation von latenten Klassen. Auch für den (tendenziell konservativeren) BIC zeigt sich ein deutlicher Effekt, welcher sich allerdings in einer Unter- und in einer Überidentifikation der latenten Klassen der Replikationen äußert. Entgegen den Erwartungen zeigt sich für gruppenspezifische Messunterschiede in den Intercepts (Invar=4) für alle IC-Indizes ein positiver Zusammenhang mit dem Grad an Schiefverteilung der Klassenanteile. Je kleiner eine Gruppe wird, desto seltener wird eine Replikation fälschlicherweise als 4-Klassenlösung identifiziert. Eine mögliche Deutung hierfür

wäre, dass durch eine geringere Anzahl von Fällen in einer Klasse auch die durch die Intercept-Unterschiede erzeugte Klassen-Heterogenität reduziert wird, was wiederum die allgemeine Tendenz zur Überidentifikation in der 4. Invarianzkonfiguration abschwächt. Für die 5. und 6. Invarianzform, welche wiederum die Unterschiede hinsichtlich der Faktorladungen ausweisen, dreht sich dieser Zusammenhang und führt in der groben Tendenz wieder zu einer Klassenüberidentifikation. Es kann festgehalten werden, dass sich entgegen den ersten Erwartungen keine einheitliche Interaktionswirkung von geringer Klassengröße und Verletzungen der Messinvarianz ausmachen lässt. Eine geringe Klassengröße von 5 % scheint zwar insgesamt einen verstärkenden negativen Effekt zu haben, aber nur bei Konfigurationen, welche sowohl Verzerrungen der Intercepts als auch der Faktorladungen aufweisen. Und selbst in diesen Konfigurationen ist der Effekt nicht über die IC-Indizes hinweg stabil. aBIC und BIC scheinen insgesamt sensibler zu reagieren, wohingegen der AIC relativ robust (wenn auch an der absoluten Klassenidentifikationsraten gemessen keineswegs optimal) gegenüber dem Zusammenspiel von Messinvarianzverzerrungen und geringer Klassengröße ist.

9.3 Modellbasierte Analyse bzgl. AIC, BIC, aBIC

Für die modellbasierte Auswertung zur Klassenidentifikation wurden drei abhängige Variablen für die IC-Indizes AIC, BIC und aBIC erstellt, welche den Wert «1» annehmen, wenn in der jeweiligen Replikation das Modell mit der korrekten Klassengröße von 3 als richtiges/bestes Modell identifiziert wurde. Wenn hingegen die 1-, 2- oder 4-Klassenlösung in der Replikation als richtig identifiziert wurde, dann nehmen die abhängigen Variablen den Wert «0» an.

Diese neu gebildeten Variablen wurden als abhängige Größen in logistischen Regressionsmodellen verwendet. Die unabhängigen Variablen der Regression sind die Simulationsgrößen und die Interaktionsterme zwischen diesen Größen. Bei der modellbasierten Auswertungen werden nicht mehr nur Teilgruppen von Simulationskonfigurationen betrachtet (wie zuvor durch das Konstanthalten der nicht betrachteten Simulationsgrößen). In die Analysen fließen alle 162 Simulationskonfigurationen mit jeweils 300 Replikationen³⁹ ein. Dies führt zu einer sehr hohen Fallzahl, weshalb es nicht mehr sinnvoll erscheint, dass die Signifikanz eines Effekts als Kriterium für dessen Interpretation herangezogen

³⁹Für 6 Simulationskonfiguration sind nicht alle 300 Replikationen konvergiert (sondern jeweils nur 299 Replikationen). Dies führt zu einer Gesamtfallzahl von 48594 für die Modellschätzungen.

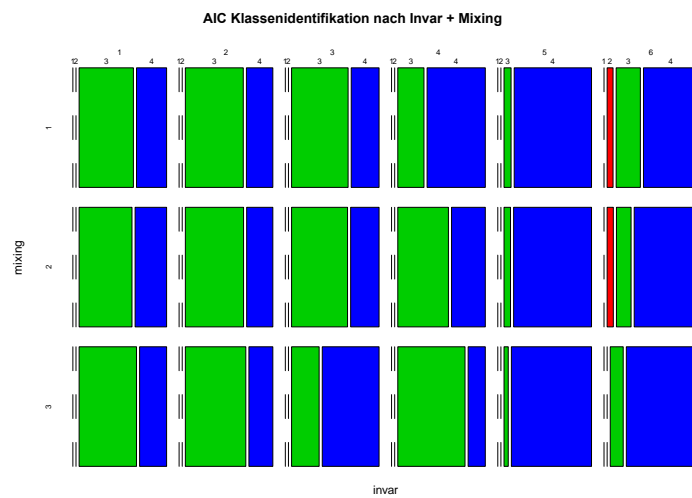
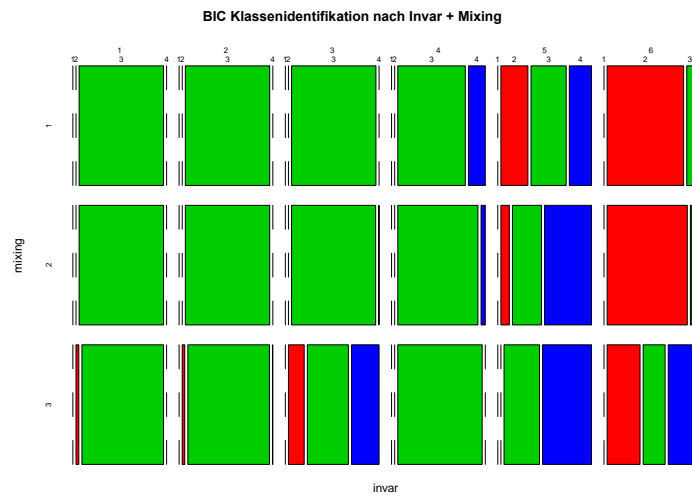
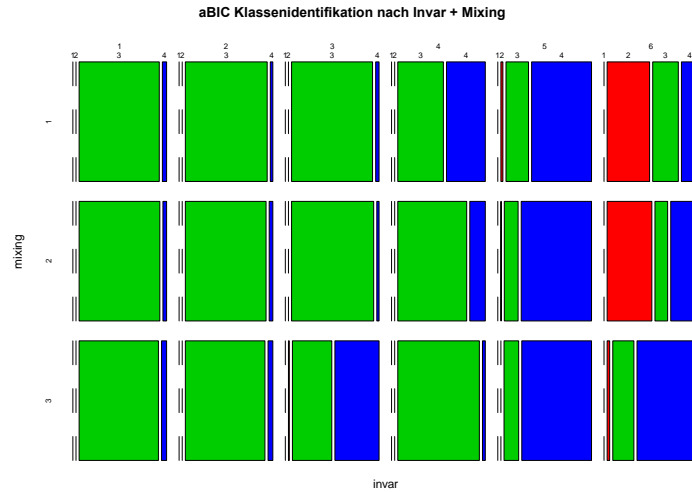


Abbildung 10: Interaktionseffekt zwischen Invarianzabstufungen und Klassenanteilen (Invar*Mixing) bzgl. Klassenidentifikation nach AIC, BIC und aBIC.

Replikationsanzahl pro Konfiguration jeweils 300.

Legende: Schwarz=1-Klasse, Rot=2-Klassen, Grün= 3-Klassen; Blau=4-Klassen

wird. Da durch den Zufallsprozess der Monte-Carlo-Simulation ein gewisses «Rauschen» in der simulierten Datenbasis erzeugt wird, dürfen diese Effekte, auch wenn sie bedingt durch die hohe Fallzahl signifikant werden, nicht überinterpretiert werden. Ein geeigneteres Kriterium für die Interpretation stellt die Effektstärke dar, welche hier anhand der Odds-Ratio-Werte bestimmt wird. Zur inhaltlichen Interpretation eines Effekts wurde eine Effektstärke von mindestens ± 0.2 d (Cohen's d) (bzw. Pearsons r von 0.1) festgelegt, was bei einer Umrechnung in Odds-Ratios (Umrechnung nach Borenstein et al. (2009)) in etwa einem Odds-Ratio-Wert von über 1.4 (bzw. unter 0.70) entspricht. Aus Abbildung 11 können die Average Marginal Effects (AME) Schätzer (mit den dazugehörigen 95%-Konfidenzintervallen) für die Haupt- und Interaktionseffekte der logistischen Regressionsmodelle⁴⁰ entnommen werden.⁴¹ Da die unabhängigen Simulationsgrößen dummykodiert in die Regressionsmodelle eingehen, sind die AME-Schätzer jeweils bezogen auf die erste Kategorie der Simulationsgröße zu interpretieren.⁴² Die AME-Schätzer geben somit die über alle Beobachtungen gemittelte durchschnittliche Änderung der Wahrscheinlichkeit für eine richtige Klassenidentifikation wieder, welche unter statistischer Kontrolle der restlichen Modellvariablen im Vergleich zur jeweiligen Referenzkategorie erzielt wird (vgl. Urban & Mayerl, 2018: 406). Die in Abbildung 11 gepunktet eingezeichneten Orientierungslinien verlaufen bei $y = \pm 0.05$ und grenzen die nicht inhaltlich interpretationsfähigen Effekte von den tatsächlich bedeutsamen Einflüssen ab. Die AME-Schätzer können zwar nicht unmittelbar in die Effektstärkegrenzen der Odds-Ratios überführt werden, aber ein Vergleich mit Tabelle 44 zeigt, dass die Grenzziehung für die hier berichteten Effekte (nahezu) deckungsgleich ist.

9.3.1 Modellbasierte Haupteffekte

Bzgl. des Haupteffekts der Messinvarianzabstufungen zeigt sich wie bereits in der deskriptiven Analyse der prozentualen Klassenidentifikationstabellen, dass längsschnittliche Verzerrungen keinen bedeutenden Einfluss auf die korrekte Klassenidentifikation haben,

⁴⁰Aufgrund von Multikollinearitätsproblemen und einer erhöhten Wahrscheinlichkeit für vollständige bzw. quasi-vollständige Separation konnte der Interaktionsterm zwischen der Messinvarianz und der Klassenseparation nicht zusätzlich in die Schätzmodelle mit aufgenommen werden. Alternative Modelle wurden jedoch geschätzt und widersprechen den Ergebnissen aus Abschnitt 9.2 inhaltlich nicht.

⁴¹Die vollständigen Modelle können zusätzlich im Anhang A aus den Tabellen 43 (Logitkoeffizienten) und 44 (Odds-Ratio) entnommen werden.

⁴²Die Referenzkategorien sind folglich: Vollständige Invarianz der Messkonstrukte über die Zeit und zwischen den Gruppen (Invar=1); eine Fallzahl von 300 (Nobs=1); eine Klassendistanz von 0.8 MD (Sep=1) und eine prozentuale Gleichverteilung der Fälle über die drei latenten Klassen (Mixing=1).

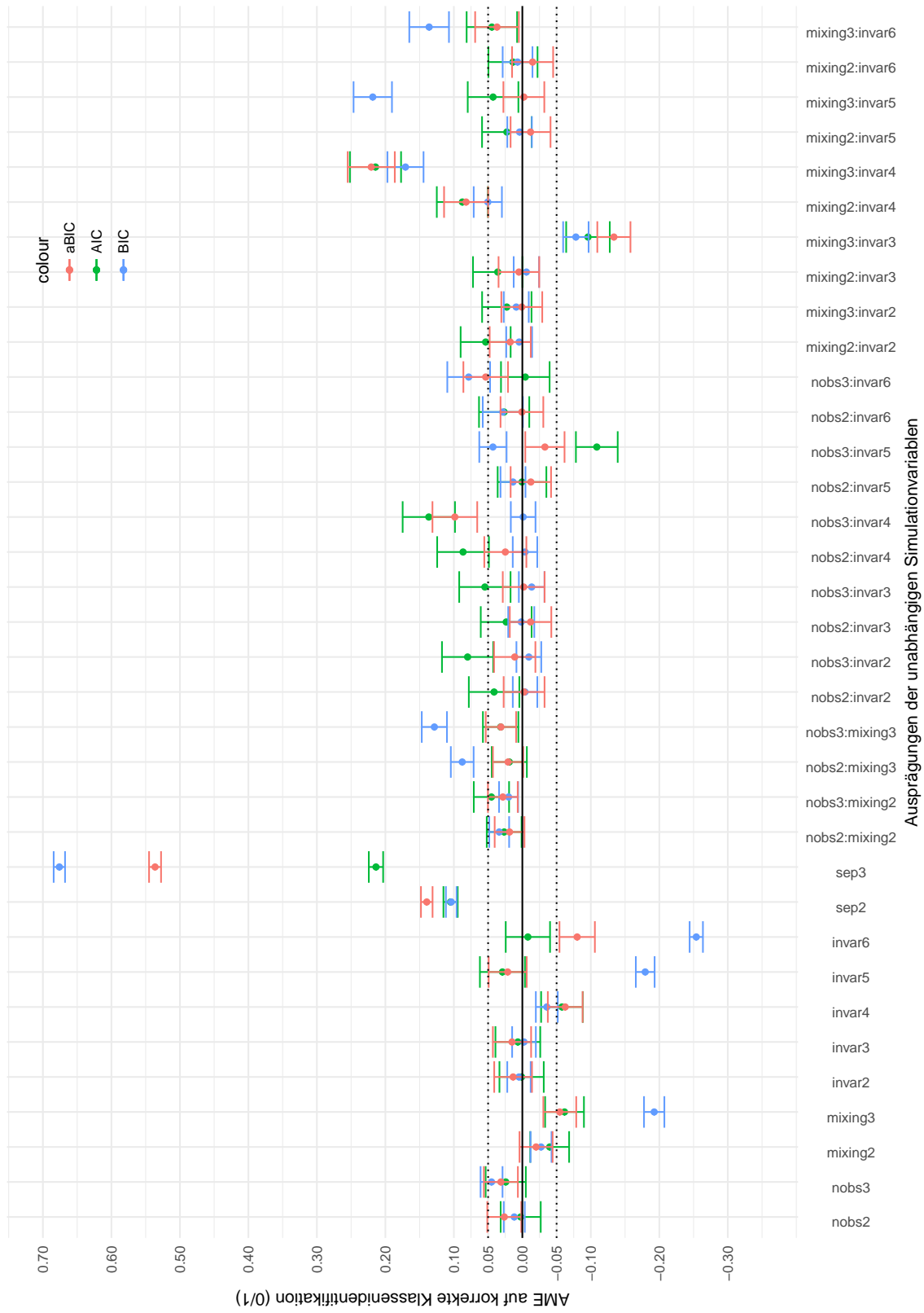


Abbildung 11: AMEs bzgl. der abhängigen Variablen der korrekten Klassenidentifikation nach AIC, BIC und aBIC

Die Fallzahl der Schätzungen beträgt jeweils 48594.

unabhängig vom zugrundeliegenden IC-Index. Für Gruppenunterschiede in den manifesten Intercepts lässt sich für den AIC und den aBIC ein, im Durchschnitt aller Beobachtungen und unter Kontrolle der restlichen Modellvariablen, leicht negativer Effekt ausmachen. Für den BIC ist der Effekt zwar gleichgerichtet, aber nicht mehr als bedeutsam einzustufen. Für Gruppenunterschiede in den Intercepts und in den Faktorladungen dreht sich das Bild. Für AIC und aBIC kann weder ein positiver noch negativer Effekt bzgl. der richtigen Klassenidentifikation im Vergleich zur Referenzkategorie ausgemacht werden. Der BIC zeigt hingegen eine um durchschnittlich 17.8 % reduzierte Wahrscheinlichkeit für eine korrekte Klassenidentifikation im Vergleich zur Referenzkategorie mit völliger Messinvarianz. Dieser Effekt verstärkt sich für die kombinierte Invarianzform noch einmal auf 25 % für den BIC. Auch der aBIC zeigt in der Konfiguration mit längsschnittlicher und gruppenspezifischer Messinvarianz mit 8 % einen bedeutsamen negativen Zusammenhang. Lediglich die Klassifikationsrate des AIC wird durch die Invarianzstufen 5 und 6 im Durchschnitt nicht beeinflusst. In der allgemeinen Tendenz gleichen sich die Haupteffekte der Invarianzabstufungen der Modelle und die der deskriptiven Tabellenanalyse. Für die fehlende Messinvarianz auf der Gruppenebene und in der Kombination von zeitlicher und gruppenspezifischer Non-Invarianz lassen sich jedoch graduelle Unterschiede ausmachen. Da die in den Modellen errechneten Effekte allerdings über alle Konfigurationen gemittelt sind, wird dadurch keineswegs ein Widerspruch zu den deskriptiven Ergebnissen hergestellt.

Insgesamt fällt auf, dass der AIC auf verschiedene Formen der Invarianz wenig sensitiv reagiert. Der aBIC und (insbesondere) der BIC reagieren bzgl. der korrekten Klassenidentifikation negativer.

Auch für den Haupteffekt der Fallzahl zeigt sich in den Modellschätzungen ein gradueller Unterschied zu den deskriptiven Analysen. Während der aBIC und (in geringerem Umfang) der AIC in den Tabellenanalysen noch mit besseren Klassifikationsraten bei einer Erhöhung der Fallzahl reagierten, zeigt sich dies im Mittel über alle Konfigurationen (und unter Kontrolle aller restlichen Modellvariablen) nicht mehr. Zwar zeigen die AME-Schätzer in die erwartete positive Richtung, was drauf hindeutet, dass eine höhere Fallzahl die korrekte Klassenidentifikation begünstigt, aber das absolute Ausmaß der Effekte übersteigt den zuvor gesetzten Grenzwert für eine substantielle Interpretation nicht. Die Schätzergebnisse zu den Anteilsverhältnissen zwischen den Klassen stimmen wiederum

deutlicher mit den deskriptiven Tabellenanalysen überein. Für die moderat schiefverteilte Ausprägung zeigen sich keine nennenswerten negativen Effekte. Die Ausprägung mit einer kleinen latenten Gruppe von 5 % beeinflusst die Klassenidentifikation für alle IC-Indizes negativ. Während der Effekt mit einer um 5 bis 6 % schlechteren Wahrscheinlichkeit für die korrekte Identifikation des AICs und des aBICs noch schwach ausfällt, fällt er für den, hinsichtlich der Klassenidentifikation konservativeren, BIC mit einer Reduktion um 19-%-Punkte deutlich stärker aus.

Auch die Einflüsse der Klassendistanz stimmen mit den vorangegangenen bi- und trivariaten Analysen überein. Wie bereits durch die Analyse der Kontingenztabellen nahegelegt wurde, geht der stärkste Einfluss des gesamten Modells von der Klassendistanz aus. Bereits eine Steigerung von 0.8 MD auf 1.5 MD geht durchschnittlich mit einer um 10 bis 14-prozentigen Steigerung der Rate der richtigen Klassenanzahl einher. Eine Erhöhung der Klassendistanz von 0.8 MD auf 4.0 MD geht darüber hinaus mit einer Steigerung um 67.8- (BIC), 53- (aBIC) und 21-%-Punkte (AIC) einher. Diese AME-Schätzer übersteigen alle Haupt- und Interaktionseffekte im Modell und stellen die Wichtigkeit der Klassendistanz für die IC-Indizes im Rahmen von MIGMMs heraus. Zieht man die Ergebnisse der deskriptiven Tabellenanalysen in Betracht, ist der geringe Einfluss auf den AIC höchstwahrscheinlich auf die durchschnittliche Überidentifikation des AIC bei 4-Klassenlösungen zurückzuführen.

9.3.2 Modellbasierte Interaktionseffekte

Fallzahl und Mixing Proportions. Bereits bei der Darlegung der Annahmen zu den Anteilsverhältnissen zwischen den Klassen wurde angenommen, dass Probleme insbesondere dann zu erwarten sind, wenn kleine Klassengrößen mit geringen Fallzahlen kombiniert werden. Die Auswertung anhand der Mosaikplots konnte diesen Zusammenhang für ansonsten «optimale» Bedingungen bestätigen. Auf Grundlage der Modelle kann diese Annahme, über alle Konfigurationen betrachtet, partiell bestätigt werden. Alle Ausprägungen des Interaktionsterms zeigen einen positiven Einfluss von höherer Fallzahl bei kleinen Klassen. Allerdings überschreiten die AME-Schätzer nur für das BIC-Modell den Grenzwert für eine substantielle Interpretation (mit knapp 13-%-Punkten). Folglich kann auch die (negative) Sensitivität des BIC bei einer kleinen Klassengröße durch eine hohe Fallzahl fast kompensiert werden. Der gleiche Zusammenhang gilt zwar rein rechnerisch auch für AIC und aBIC, aber weder die Haupteffekte noch die Interaktionsterme sind

hierbei stark genug, um inhaltlich relevant zu sein. Entgegengesetzt betrachtet, hat eine höhere Fallzahl zwar durchschnittlich keinen deutlichen Haupteffekt auf die richtige Klassenidentifikation, kann aber bei geringen Klassengrößen kompensierend wirken und die korrekte Identifikation begünstigen. Folglich sind höhere Fallzahlen, unter der Bedingung, dass kleinere Gruppen in der Stichprobe vermutet werden, generell zu empfehlen.

Messinvarianz und Fallzahl. Die moderierende Wirkung der Fallzahl auf Invarianzverletzungen (siehe hierzu auch Abschnitt 9.2) lässt sich auch anhand der Modellschätzungen nicht gänzlich bestätigen und unterscheidet sich für die einzelnen IC-Indizes. Der BIC zeigt sich sehr robust gegenüber einer vermeintlich verstärkten Wirkung von geringer Fallzahl auf eine falsche Spezifikation der Messmodelle. Lediglich für die Fallzahlausprägung von 1300 zeigt sich in der Kombination aus gruppen- und längsschnittlicher Invarianzverletzung ein leicht positiver Interaktionseffekt. Dieser kann allerdings nicht einmal im Ansatz den negativen Haupteffekt ausgleichen.

Der AIC reagiert stärker auf die moderierende Wirkung der Fallzahl. Für die ersten vier Invarianzstufen lässt sich insbesondere bei der höchsten Fallzahl eine Tendenz zu besseren korrekten Identifikationen ausmachen. Sobald jedoch in Invarianz-Kategorie 5 Gruppenunterschiede in den Faktorladungen eingeführt werden, dreht sich der Zusammenhang und geht mit weniger korrekten Identifikationen einher. Für die kombinierte Invarianzabstufung lässt sich kein moderierender Effekt mehr entdecken. Der aBIC liegt bzgl. der Sensitivität dieser Interaktionsbeziehungen zwischen dem BIC und dem AIC. Eine verstärkende Wirkung ist nur für die gruppenspezifischen Messunterschiede in den Intercepts festzustellen. Zusammenfassend kann von keiner einheitlichen Verstärkung des Effekts der Invarianzverletzungen durch geringe Fallzahlen gesprochen werden. Lediglich der Invarianzeffekt des AICs wird durchschnittlich über die Simulationskonfigurationen hinweg moderiert. Falls folglich Zweifel bzgl. der zeitlichen Stabilität der Messung bestehen, kann für den AIC eine Erhöhung der Fallzahl angeraten werden. Sobald jedoch Zweifel an der Äquivalenz der Messung zwischen den latenten Klassen besteht, kann dies nicht mehr unbedingt angeraten werden. Für den BIC und (in geringerem Ausmaß) den aBIC können verzerrende Invarianzverletzungen nicht wesentlich kompensiert (aber auch nicht zugespitzt) werden.

Messinvarianz und Mixing. Wie sich bereits bei der deskriptiven Auswertung der Kontingenztabellen zeigte, besteht für längsschnittliche Invarianzverletzungen erst dann ein Interaktionseffekt, wenn sowohl Intercepts wie auch Faktorladungen verzerrt sind. Für die dritte Invarianzkategorie zeigt sich bei einer kleinen Klasse, dass die richtige Klassenidentifikation für alle IC-Indizes erschwert wird.

Für Messunterschiede zwischen den latenten Gruppen dreht sich diese Beziehung. Eine besonders kleine Klasse geht hier, in der Tendenz, mit einer besseren Identifikationsrate einher. Für Interceptunterschiede (Invar=4) ist diese verstärkende Wirkung besonders deutlich. Im Vergleich zur Gleichverteilung (und Messinvarianz) verbessert sich die Klassenidentifikation für die drei IC-Indizes zwischen 17 und 22 %-Punkten. Wiederum scheint die Deutung plausibel, dass durch kleinere Klasse zusätzlich Heterogenität zwischen den Klassen (bedingt durch Interceptverzerrungen) aus den Daten herausgenommen wird. Für den AIC und den aBIC verschwindet dieser positive Effekt in den Invarianzstufen 5 und 6 allerdings wieder. Für den BIC bleibt der positive Interaktionseffekt von kleinen Klassen in den Invarianzstufen 5 und 6 erhalten (mit 14 bis 21 %-Punkten). Der besonders negative Haupteffekt der Invarianzstufen 5 und 6 auf den BIC kann hier durch den Moderationseffekt der Klassengröße kompensiert werden.

Alles in allem muss festgehalten werden, dass eine Interaktionsbeziehung zwischen Messinvarianzverletzungen und kleinen Klassengrößen besteht. Da die Richtung und Wirkung dieses Effekts allerdings stark vom IC-Index und der jeweiligen Datensituation abhängen, lassen sich hieraus keine allgemeingültigen Empfehlungen ableiten. Während kleine Klassen bei längsschnittlichen Invarianzverletzungen ein Problem darstellen können, gibt es für gruppenspezifische Messunterschiede Situationen, in welchen sie die korrekte Klassenidentifikation begünstigen können.

9.4 Diskussion zur IC-Klassenidentifikation

Dass GMM- und SOGMM-Schätzungen weder zu viele noch zu wenige latente Klassen in den Daten finden, ist vielleicht das zentrale Kriterium für ihre Performanz und wichtige Bedingung für die sinnvolle Anwendung in der empirischen Forschungspraxis. Schließlich stellt die (oft) explorativ betriebene Post-hoc-Klassenidentifikation den genuinen Anwendungsfall für diese Klasse von Modellen dar. Wie die hier durchgeführten Auswertungen zeigen, kann die Bedingung der richtigen Klassenauffindung nicht für alle Datensituationen befriedigend eingehalten werden. Sowohl die Datensituation (bzw. Simulationskonfigurati-

on) als auch der verwendete IC-Index haben einen Einfluss darauf, ob die Klassenstruktur der längsschnittlichen Entwicklungsverlaufsdaten im Rahmen von MIGMM-Schätzungen zuverlässig aufgedeckt werden kann.

In den Auswertungen ist zunächst positiv aufgefallen, dass für keinen der IC-Indizes längsschnittliche Verletzungen der Äquivalenz der Messung einen nennenswerten Effekt auf die Klassenidentifikation ausüben. Dies legt die Aussage nah, dass eine Veränderung der Messung über die Zeit die Leistungsfähigkeit von MIGMMs zur korrekten Klassenidentifikation nicht beeinflusst und daher (prinzipiell) vernachlässigt werden kann, wenn man nur an der Gruppenanzahl interessiert sein sollte. Allerdings gelten hierfür zwei wichtige Einschränkungen. (1) Für Konfigurationen mit kleinen Klassen konnte gezeigt werden, dass fehlende Messinvarianz über die Zeit zum Problem werden kann und die Wahrscheinlichkeit für eine korrekte Klassenidentifikation verringert. (2) Darüber hinaus zeigt die Auswertung, dass sich in Kombination mit gruppenspezifischen Unterschieden auch die längsschnittliche Non-Invarianz der Messung auf die Klassenidentifikation nach den IC-Indizes auswirkt. Folglich sollte die Invarianz der Messung über den Erhebungszeitraum hinweg aufgrund dieser Interaktionseffekte dennoch sichergestellt werden um eine zuverlässige Klassenidentifikation sicherzustellen. Der längsschnittliche Invarianztest ist im Rahmen einer konfirmatorischen Faktorenanalyse möglich, sofern latente Messmodelle eingesetzt werden. Somit empfiehlt sich bereits bei längsschnittlichen Messungleichheiten die Verwendung latenter Messmodelle im Rahmen von Growth Mixture Modellierungen statt einfacher Composite Score Modelle (sprich MIGMMs statt GMMs), um die korrekte Klassenidentifikation zu begünstigen.

Für Messunterschiede zwischen den latenten Gruppen fällt das Ergebnis noch deutlicher aus. So zeigt sich hinsichtlich der Klassenidentifikation bei Gruppenunterschieden in der Tendenz ein deutlich schlechteres Ergebnis als bei den längsschnittlichen Non-Invarianzen. Gleiches gilt auch für die Kombination aus längsschnittlichen und gruppenspezifischen Messunterschieden. Besonders der BIC, aber auch der aBIC, zeigten sich hierfür sensitiv. Auch wenn sich diese Haupteffekte durch höhere Fallzahlen teilweise kompensieren lassen, wird hierdurch der Haupteffekt nicht aufgehoben. Auch kompensierende Effekte von kleinen Klassen können das Ergebnis der Haupteffekte nicht gänzlich aufheben.⁴³ Dieses Ergebnis ist insofern problematisch, als dass Messgleichheit zwischen den latenten

⁴³Lediglich die besonders schlechte Leistung des BIC bei Unterschieden in den Faktorladungen in den Invarianzkonfigurationen 5 und 6 kann abgefangen werden.

Klassen nicht vor der Schätzung von MIGMMs sichergestellt werden kann, wie bei der längsschnittlichen Messung. Da die Anzahl und die Zuordnung der Klassen erst durch das Schätzverfahren selbst hergestellt werden, kann keine Messmodellüberprüfung vor der Schätzung stattfinden. Auch eine Post-hoc-Prüfung ist nicht sehr erfolgversprechend, da die Zusammensetzung der latenten Klassen bereits verzerrt sein kann und nur deshalb eine vermeidliche Un-/Gleichheit anzeigt.⁴⁴ Allerdings sollte sich die empirische Praxis nicht auf diese fehlende Testbarkeit zurückziehen und das Problem gänzlich ignorieren. So kann zum einen angeraten werden, dass sichergestellt wird, dass eine längsschnittliche Invarianz besteht, um zumindest die Interaktion zwischen gruppenspezifischer und längsschnittlicher Invarianz weniger wahrscheinlich zu machen. Darüber hinaus scheint es ratsam, dass für MIGMM-Schätzungen vorrangig etablierte Messkonstrukte verwendet werden, welche vorzugsweise bereits anhand verschiedener bekannter Gruppen hinsichtlich Validität und Reliabilität überprüft wurden. Dadurch kann wenigstens die Wahrscheinlichkeit dafür erhöht werden, dass die Invarianz der Messung auch zwischen den latenten – vorher nicht bekannten – Gruppen Bestand hat. Wenn dies nicht möglich ist, so sollte man Ergebnisse zur Klassenidentifikation mit Vorsicht interpretieren, sich des Problems zumindest bewusst sein und bei der Ergebniseinordnung berücksichtigen.

Auch wenn eine höhere Fallzahl unter ansonsten optimalen Simulationsbedingungen einen positiven Einfluss auf die richtige Klassenidentifikation hat, ist der über alle Replikationen und Konfigurationen berechnete Haupteffekt so schwach, dass er nicht mehr substantiell interpretiert werden sollte, auch wenn die Richtung der Effekte den Erwartungen entspricht. Insgesamt konnten mit den Ausprägungen von 300, 700 und 1300 Fällen vermutlich keine kritischen Schwellenwerte überschritten werden.⁴⁵ Zumindest die Interaktion mit der kleinen Klassengröße liefert einen Hinweis darauf, dass eine geringere Fallzahl zu deutlicheren Problemen hätte führen können. Die Interaktion mit den Klassenanteilsverhältnissen macht darüber hinaus deutlich, dass Fallzahleinschätzungen für Mixture-Modellierungen natürlich immer in Bezug zur Anzahl der latenten Klassen betrachtet werden müssen und nicht absolut.

Sowohl in den Tabellenanalysen als auch in der modellbasierten Auswertung konnte die

⁴⁴Tatsächlich lassen sich Messunterschiede zwischen latenten Klassen bei MIGMM-Schätzungen nur im Rahmen von MC-Studien nachweisen (durch Populationsdefinition), da dies anhand empirischer Daten aufgrund der aufgeführten Gründe prinzipiell nicht möglich ist.

⁴⁵Auch die Anzahl der Messzeitpunkte kann als Informationsersatz, neben der tatsächlichen Fallzahl, relevant sein. Da die Messzeitpunkte allerdings nicht variiert wurden, bleibt dies spekulativ und müsste gesondert getestet werden.

Klassendistanz als wichtigste Bedingung für eine korrekte Klassenidentifikation herausgestellt werden. Die Bedeutung und Richtung der Beziehung sind dabei nicht überraschend und wurden auch durch die bisherigen Ergebnisse im Rahmen von Simulationsstudien zu GMMs häufig bestätigt. Überraschend ist allerdings das Ausmaß der Bedeutung der Klassendistanz. Bisherige Simulationsstudien legten den Schluss nahe, dass Werte von 1.5 MD als mittlere oder große Klassendistanz einzustufen sind und daher auch gute Identifikationsraten aufweisen. Für MIGMM-Schätzungen kann diese Einstufung, gemessen an den Ergebnissen der durchgeführten Simulation, nicht gelten. Erst ab einer Distanz von 4.0 MD konnten akzeptable Identifikationsraten beobachtet werden. Natürlich kann angemerkt werden, dass die Bandbreite zwischen 1.5 und 4.0 MD sehr weit ist und schon vor der sehr hohen Klassendistanz von 4.0 akzeptable Identifikationsraten erkennbar wären. Dies ändert allerdings nichts daran, dass die üblichen Grenzwerte von GMMs für MIGMM nicht gelten können. Vielleicht wäre der Effekt auch nicht so deutlich ausgefallen, wenn für die Simulation ein anderes Verfahren gewählt worden wäre, um die Distanz der Klassen herzustellen. In der Simulations-Literatur gibt es immer noch kein Konsens darüber, was die adäquate Methode darstellt, um Klassendistanz herzustellen. So wird, beispielsweise, in einigen Untersuchungen die Klassendistanz lediglich über Unterschiede der Intercepts hergestellt. Allerdings erscheint die hier gewählte Methodik über Unterschiede in den Intercept- und den Slope-Termen die näher an der Empirie liegende Option zu sein. Es bleibt festzuhalten, dass die Klassendistanz, unabhängig vom verwendeten IC-Index, elementar ist, um für MIGMM-Schätzungen korrekte Klassenidentifikationen sicherzustellen. Wo der Wendepunkt zwischen 1.5 und 4.0 MD liegt, muss allerdings in weiteren Analysen untersucht werden.

Wie erwartet wurde, haben moderate Schiefverteilungen der Klassenanteile keinen nennenswerten Effekt auf die Klassenidentifikation; unabhängig vom verwendeten IC-Index. Erst besonders kleine Klassen führen zu Problemen. Der BIC zeigt sich hierfür sowohl nach den modellbasierten Auswertungen wie auch nach den deskriptiven Kontingenztabellenanalysen besonders sensibel. Wie die Interaktionsbeziehung mit der Fallzahl zeigt, können diese Probleme allerdings, zumindest deutlich für den BIC, mit mehr Beobachtungsfällen innerhalb der Klassen kompensiert werden. So betrachtet kann von keinem eigenständigen Effekt der Klassenverhältnisse gesprochen werden. Vielmehr scheint die fehlende Beobachtungsinformation in der Datenbasis das Problem zu sein. Insgesamt können allerdings die Daumenregeln, welche in der Literatur bereits für GMM-Schätzungen aufgestellt wurden,

bestätigt werden. Die üblicherweise vorgeschlagenen Grenzwerte von 5 oder 7 % scheinen auch im Rahmen von MIGM-Modellen ihre Berechtigung zu haben.

In der Forschungspraxis stellt sich die Frage, auf welchen IC-Index man sich nun im Rahmen von MIGMM-Schätzungen stützen sollte. Leider lassen sich anhand der Simulationen keine einfachen Empfehlungen für ein Maß ableiten, welches in jeder Situation die besten Resultate liefert. Anhand der Analysemodelle lässt sich sagen, dass die Klassenidentifikation des BIC am stärksten durch die simulierten Datensituationen beeinflusst wird ($Nagelkerke - R^2 = 0.72$). Der AIC verhält sich hinsichtlich der Simulationsgrößen am robustesten ($Nagelkerke - R^2 = 0.074$)⁴⁶, während der aBIC ($Nagelkerke - R^2 = 0.32$) zwischen dem BIC und dem AIC liegt. Nun darf die Beeinflussbarkeit oder Robustheit eines IC-Indexes allerdings nicht als alleiniges Kriterium für dessen Performanz herangezogen werden. Auch das Grundniveau der Klassenidentifikation muss berücksichtigt werden, um eine informierte Auswahl zwischen den IC-Indizes zu treffen. So zeigt der AIC unter ansonsten optimalen Bedingungen eine deutliche Tendenz zur Überidentifikation von latenten Klassen, wobei er z. B. nur in 64% der Fälle die 3-Klassenlösung findet. BIC und aBIC zeigen unter optimalen Bedingungen mit 100 % und 95% deutlich bessere Grundniveaus der Klassenidentifikation. Allerdings sind diese guten Ergebnisse dann wiederum stärker von den jeweiligen Datenbedingungen abhängig. Auf Grundlage der durchgeführten Simulationen können folgende vorläufige Anwendungsempfehlungen für die IC-Maße abgeleitet werden.

(1) Der BIC eignet sich als Entscheidungskriterium vor allem für Situationen, in welchen zuverlässige Messkonstrukte Anwendung finden, ein klarer und starker Klassenunterschied bzgl. der Verlaufsmuster auf Grundlage von theoretischen, analytischen oder empirischen Vorüberlegungen besteht und keine der Klassen besonders klein ausfällt. Die Fallzahl sollte, solange sie sich im Rahmen der Simulationsausprägung bewegt, keine entscheidende Rolle spielen. Unter diesen Bedingungen wird der BIC sehr zuverlässige Ergebnisse bzgl. der Klassenidentifikation aufweisen und ist den anderen IC-Maßen überlegen. (2) Sollte in der empirischen Praxis keine hinreichende Information zur Messqualität des latenten Konstrukts, der Klassendistanz oder für die Existenz von kleinen Gruppen vorhanden sein (was wohl eher der Regelfall als die Ausnahme ist), so sollte vorrangig der aBIC als

⁴⁶Diese Robustheit des AICs gegenüber den Simulationsgrößen zeigte sich auch bei Kim & Wang (2017) (Kim & Wang, 2017: 15).

Kriterium für die Klassenidentifikation herangezogen werden. Diese Empfehlung basiert darauf, dass bei der Auswertung gezeigt werden konnte, dass er sowohl unter optimalen Datenbedingungen gute Identifikationsraten aufweist sowie auch bei problematischen Datensituationen weniger sensibel auf fehlende Messinvarianz, kleine Gruppen und geringe Klassendistanz reagiert. (3) Darüber hinaus kann der AIC trotz seiner schlechten Performanz bzgl. des Grundniveaus bei der Klassenidentifikationen bei MIGMM-Schätzungen als Kennwert für die obere Schranke der Klassenanzahl hinzugezogen werden. Sollte der AIC in suboptimalen Datensituationen eine höhere Klassenanzahl nahelegen, sollte diese Möglichkeit bei der Ergebnisdeutung beachtet werden. Natürlich gelten diese, als vorläufig einzustufenden, Empfehlungen für die IC-Maße im Rahmen von MIGMMs strenggenommen nur für Modellschätzungen, welche die Charakteristika der durchgeführten Simulationsstudie teilen und können nicht ohne Weiteres auf alternative Spezifikationen oder Datensituationen ausgeweitet werden. Ein zusammenfassender Überblick der diskutierten Ergebnisse kann auch aus Tabelle 10 entnommen werden.

Tabelle 10: Zusammenfassung der IC-Performanz bzgl. der korrekten Klassenidentifikation

	AIC	BIC	aBIC
Grundniveau	-	++	+
Fehlende Messinvarianz	unbedeutend; für hohe Fallzahlen pos.	unbedeutend	unbedeutend
	unbedeutend	neg. beeinflusst	unbedeutend
	unbedeutend	stark neg. beeinflusst	neg. beeinflusst
Hohe Klassendistanz	pos. beeinflusst	stark. pos. beeinflusst	stark pos. beeinflusst
Hohe Fallzahl	unbedeutend	unbedeutend	unbedeutend
Kleine Klasse	unbedeutend	neg. beeinflusst	unbedeutend
Einsatz-Empfehlung	Hinweis für maximale Klassenanzahl bei suboptimalen oder unklaren Datensituationen	Einscheidungskriterium bei optimalen Datensituationen (Messinvarianz, Klassen >5%, große Klassendistanz)	Entscheidungskriterium bei suboptimalen oder unklaren Datensituationen

9.5 Zusatzsimulation: Ergebnisse bzgl. BLRT

Der BLRT folgt mit seinem auf Resampling basierendem Likelihood-Ratio-Test einer anderen – und rechentechnisch aufwendigeren – Modellselektionslogik wie die IC-Maße. Die Ergebnisse zur Performanz des BLRT werden im Weiteren zwar getrennt dargestellt, aber dennoch mit der Performanz der IC-Maßzahlen in Beziehung gesetzt. Bei den folgenden berichteten Daten handelt es sich strenggenommen um das Ergebnis einer separaten Simulation, welche nicht direkt mit den Simulationen zu den IC-Maßen verglichen werden kann (vgl. hierzu die Ausführungen in Abschnitt 7.3.5). Grund hierfür ist, dass die Anzahl der Replikationen und die randomisierten Startwertesets aus praktischen Erwägungen anders als in der Hauptsimulation festgelegt werden mussten, wodurch sich eine (in Teilen) andere Datengrundlage ergibt. Die Ergebnisse für den BLRT basieren folglich aus forschungspragmatischen Gründen lediglich auf 200 Replikationen, welche jeweils mit 120 randomisierten Startwertesets geschätzt wurden.⁴⁷ Dies hat zur Folge, dass die Ergebnisse auch mit einem höheren Grad an Unsicherheit behaftet sind und kleine Differenzen zwischen den Konfigurationen nicht überinterpretiert werden sollten. Die Anzahl der Bootstrapziehungen innerhalb der Modelle der einzelnen Replikationen wird über die Defaultoption von Mplus unter Verwendung eines sequentiellen Verfahrens bestimmt, bei dem die Anzahl der Züge von 2 bis 100 variiert (vgl. Muthén & Muthén, 2015: 621). Wenn der p-Wert nahe 5 % liegt, verwendet dieses sequentielle Verfahren mehr Ziehungen als wenn der p-Wert weit von 5 % entfernt ist (vgl. Nylund et al., 2007: 551 f.).⁴⁸ Die Festlegung des durch den Test identifizierten Klassenmodells innerhalb jeder Replikation kann hierbei nicht wie bei den IC-Indizes durch die kleinste Maßzahl geschehen. Innerhalb jeder Replikation wurde sequenziell bestimmt, ob das jeweils größere Klassenmodell eine signifikant bessere Passung aufweist als das nächst kleinere Modell. Diese Sequenz vom 1- bis zum 4-Klassenmodell wurde jeweils abgebrochen, sobald keine bessere Passung durch den BLRT mehr festgestellt werden konnte oder das 4-Klassenmodell als besser als das (korrekte) 3-Klassenmodell identifiziert wurde.

⁴⁷Trotz dieser Reduktion des Rechenaufwands hätten alleine die Simulationen zum BLRT auf einem handelsüblichen PC (ca. 2,6 GHz) – also ohne Nutzung des BwUniClusters – 17062.14 Stunden oder knapp 711 Tage in Anspruch genommen. Dabei wurden mit 162 Simulationskonfigurationen und 200 Replikationen bei 3 Modellklassenvarianten 97 200 Modellschätzungen durchgeführt, wobei lediglich für eine Modellschätzung keine Modellkonvergenz erreicht wurde.

⁴⁸Für eine detaillierte Beschreibung siehe auch den Appendix von Nylund et al. (2007).

9.5.1 Haupteffekte bzgl. BLRT

Zunächst werden wiederum die Effekte der einzelnen Simulationsgrößen auf die Performanz des BLRT zur Klassenidentifikation, unter ansonsten optimalen Simulationsbedingungen, dargestellt. Dafür werden die aktuell nicht betrachteten Simulationsgrößen jeweils auf folgende Größen fixiert:

- Fallzahl = 1300 Beobachtungsfälle;
- Invarianzstufe = vollständige Messinvarianz;
- Klassendistanz = 4.0 MD;
- Klassenanteilsverhältnisse = gleichverteilt über die drei Klassen.

Tabelle 11: Proz. BLRT-Klassenidentifikation nach Invarianz-Stufen

invar	1.BLRT	2.BLRT	3.BLRT	4.BLRT
1.00	0.00	0.00	93.50	6.50
2.00	0.00	0.00	86.50	13.50
3.00	0.00	0.00	89.50	10.50
4.00	0.00	0.00	84.50	15.50
5.00	0.00	0.00	10.00	90.00
6.00	0.00	21.50	47.50	31.00
Info: Sep=4 MD, Mixing=33.33.33, Nobs=1300. Fallzahl pro Konfiguration: 200 (Replikationen)				

Invarianzstufen (invar). Hinsichtlich des Grundniveaus der Klassenidentifikation bei gänzlicher Messinvarianz zeigt der BLRT ähnliche Ergebnisse wie der aBIC. Keine perfekte Identifikationsrate wie der BIC, aber zumindest ein Ergebnis nahe 95 % korrekter Klassenidentifikation. Im Vergleich zu den IC-Maßen reagiert der BLRT bereits für die beiden längsschnittlichen Nicht-Invarianzstufen mit schlechteren Identifikationsraten. Der Grad der Überidentifikation von zu vielen latenten Klassen steigt immerhin um 4 bis 7 %-Punkte an (vgl. Tabelle 11).

Bzgl. der Messunterschiede zwischen den latenten Gruppen zeigt sich ein zwiespältiges Bild. Werden lediglich die manifesten Intercepts als nicht invariant spezifiziert (invar=4), verschlechtert sich zwar die Identifikationsrate um 5 %-Punkte, aber im Vergleich zu den drei IC-Maßzahlen zeigt der BLRT hier noch das vergleichsweise beste Ergebnis. Für die Kombination aus non-invarianten manifesten Intercepts und den Faktorladungen (invar=5) dreht sich dieses Ergebnis allerdings. Hier zeigt der BLRT mit 10 % der

Replikationen eine schlechtere Identifikationsrate als der aBIC und der BIC und ist hinsichtlich der absoluten Identifikationsrate nahe am AIC. Werden längsschnittliche und gruppenspezifische Non-Invarianz kombiniert, weist der BLRT allerdings vergleichsweise gute Identifikationsfähigkeit auf. Mit gut 47 % richtiger Klassenauswahl schlägt er in dieser Konfiguration sowohl AIC, aBIC als auch BIC. In der Gesamtbetrachtung zeigt sich allerdings, dass der BLRT immer noch deutlich auf die Non-Invarianzformen reagiert und nicht robust gegenüber fehlender Messinvarianz ist. Im Vergleich zu den IC-Maßen trifft dies schon für die längsschnittlichen Konfigurationen zu. Zwar zeigt der BLRT etwas bessere Ergebnisse für die Invar-Abstufungen 4 und 6, was aber durch die besonders schlechte Performanz für die fünfte Invarianzstufe konterkariert wird.

Tabelle 12: Proz. BLRT-Klassenidentifikation nach Sep-Stufen

sep	1.BLRT	2.BLRT	3.BLRT	4.BLRT
1.00	0.00	95.50	3.00	1.50
2.00	0.00	68.50	23.00	8.50
3.00	0.00	0.00	93.50	6.50
Info: Invar=1, Mixing=33.33.33, Nobs=1300.				
Fallzahl pro Konfiguration: 200 (Replikationen)				

Klassendistanz (sep). Auch für den BLRT zeigt sich, dass die Klassendistanz die korrekte Klassenidentifikation stark beeinflusst und dies gilt auch unter ansonsten optimalen Datenbedingungen. Für hohe Klassendistanz identifiziert der BLRT minimal schlechter als der aBIC (und damit auch BIC). Für die mittlere Klassendistanz von 1.5 MD weist der BLRT eine vergleichsweise gute Identifikationsrate von 23 % auf (vgl. Tabelle 12), was natürlich absolut betrachtet immer noch keine gute Quote darstellt. Damit zeigt der Test allerdings bessere Leistung als aBIC und BIC. Lediglich der AIC zeigt mit seiner Tendenz zur Überidentifikation für geringe Klassendistanzen eine noch bessere Identifikationsrate.

Tabelle 13: Proz. BLRT-Klassenidentifikation nach Nobs-Stufen

nobs	1.BLRT	2.BLRT	3.BLRT	4.BLRT
1.00	0.00	0.50	95.00	4.50
2.00	0.00	0.00	91.00	9.00
3.00	0.00	0.00	93.50	6.50
Info: Invar=1, Mixing=33.33.33, Sep=4 MD.				
Fallzahl pro Konfiguration: 200 (Replikationen)				

Fallzahl (nobs). Die Fallzahl scheint zumindest unter optimalen Bedingungen keinen nennenswerten Effekt auf die korrekte Klassenidentifikation des BLRT zu haben. Die

Unterschiede bewegen sich maximal im Bereich von 4 % und sollten aufgrund der geringeren Replikationszahl (im Vergleich zu den Auswertungen der IC-Maße) auch nicht überinterpretiert werden (vgl. Tabelle 13). Damit ist die Leistung des BLRT bezüglich der Fallzahl zwischen aBIC und BIC einzuordnen, wobei der AIC noch die absolut schlechtesten Identifikationsraten aufweist.

Tabelle 14: Proz. BLRT-Klassenidentifikation nach Mixing-Stufen

mixing	1.BLRT	2.BLRT	3.BLRT	4.BLRT
1.00	0.00	0.00	93.50	6.50
2.00	0.00	0.00	96.50	3.50
3.00	0.00	0.50	90.50	9.00
Info: Invar=1, Nobs=1300, Sep=4 MD.				
Fallzahl pro Konfiguration: 200 (Replikationen)				

Mixing Proportions (mixing). Für die Klassenanteilsverhältnisse zeigt der BLRT für die Konfiguration mit einer sehr kleinen Klasse einen gewissen Abfall der korrekten Identifikationsrate, ähnlich wie der BIC (aber auf einem anderen absoluten Niveau). In der Gesamtbetrachtung schneidet er bzgl. der absoluten Identifikationsraten etwas schlechter als der aBIC ab. Die Unterschiede sind allerdings sowohl hinsichtlich des Vergleichs mit den IC-Maßen als auch zwischen den Konfigurationen als nicht sehr deutlich einzustufen, insbesondere wenn man die geringere Replikationszahl berücksichtigt.

Insgesamt zeigt diese erste deskriptive Auswertung, dass der BLRT hinsichtlich der Haupteffekte der Simulationsgrößen keine Sonderstellung einnimmt und keineswegs als besseres Maß eingestuft werden kann, sondern sich hinsichtlich der absoluten Performanz der Klassenidentifikation alles in allem zwischen den IC-Indizes zu verorten ist. So liegt die Leistungsfähigkeit meist zwischen aBIC und BIC, wobei z. B. für geringe Klassendistanzen auch eine Annäherung an das AIC-Maß stattfindet.

9.5.2 Modellbasierte Analysen bzgl. BLRT

Für die modellbasierte Auswertung wurde wiederum eine dichotome Variable erstellt, welche den Wert «1» annimmt, wenn innerhalb einer Replikation die korrekte Klassenanzahl vom BLRT als Lösung identifiziert wurde. Aus Abbildung 12 können die Average Marginal Effects (AME) Schätzer (mit den dazugehörigen 95-%-Konfidenzintervallen) für die

Haupt- und Interaktionseffekte des logistischen Regressionsmodells entnommen werden.⁴⁹ Die unabhängigen Simulationsgrößen im Modell sind als AME-Schätzer jeweils bezogen auf die erste Kategorie der Simulationsgröße zu interpretieren. AME-Schätzer unter 5 % werden in der weiteren Interpretation als nicht substantiell bedeutsam behandelt (analog zu den Auswertungen zu AIC, BIC, aBIC). Die Signifikanz der Effekte wird aufgrund der Fallzahl von 32399 nur sekundär betrachtet und nur als absolutes Ausschlusskriterium genutzt. Die unabhängigen Variablen der im Weiteren diskutierten log. Regression setzen sich aus den Haupteffekten der Simulationsgrößen und den Interaktionstermen zwischen diesen Größen zusammen, welche einen substantiellen Beitrag zur Modellpassung beitragen konnten.

Fallzahl. Im Durchschnitt aller Modellkonfigurationen und unter Kontrolle der restlichen Simulationsgrößen zeigt sich, dass eine hohe Fallzahl zumindest für geringe Klassendistanzen zu einer besseren Identifikationsrate führt. Dabei handelt es sich um eine um 14 und 16 %-Punkte bessere absolute Identifikation im Vergleich zur niedrigsten Fallzahlausprägung (vgl. Abbildung 12). Für eine mittlere Klassendistanz reduziert sich die Identifikationsrate (jeweils ca. um 4 %-Punkte) allerdings und verschwindet für die größte Fallzahlausprägung völlig. Betrachtet man die Interaktionsterme zwischen der Fallzahl und der Klassendistanz so wird ersichtlich, dass für hohe Klassendistanzen eine höhere Fallzahl in der durchschnittlichen Tendenz sogar zu einer schlechteren Identifikation beiträgt indem die 4-Klassenlösung⁵⁰ häufiger gewählt wird. Diese Tendenz war zwar bereits in der deskriptiven Auswertung unter optimalen Randbedingungen erkennbar, zeigt sich aber erst im Durchschnitt über alle Konfigurationen als substantiell relevant.

Mixing. Über alle Simulationsausprägungen hinweg haben kleine Klassen oder ungleichverteilte Klassenanteile im Durchschnitt keinen nennenswerten verzerrenden Einfluss auf die Performanz des BLRT. Die deskriptiven Ergebnisse unter optimalen Simulationsbedingungen können somit auch im Schnitt aller Simulationskonfigurationen bestätigt werden.

Klassendistanz und Invarianzstufen. Die Klassendistanz zeigt wiederum vergleichsweise stärksten Effekt. Bereits die mittlere Klassendistanzausprägung zeigt eine um 29

⁴⁹Die Modellschätzung kann zusätzlich im Anhang A in den Tabellen 45 (Logitkoeffizienten) und 46 (Odds-Ratio) gefunden werden.

⁵⁰Die Tendenz zur Identifikation von zu vielen Klassen ergibt sich nicht aus dem Metamodell selbst, sondern erst in Kombination mit der Betrachtung der deskriptiven Häufigkeiten.

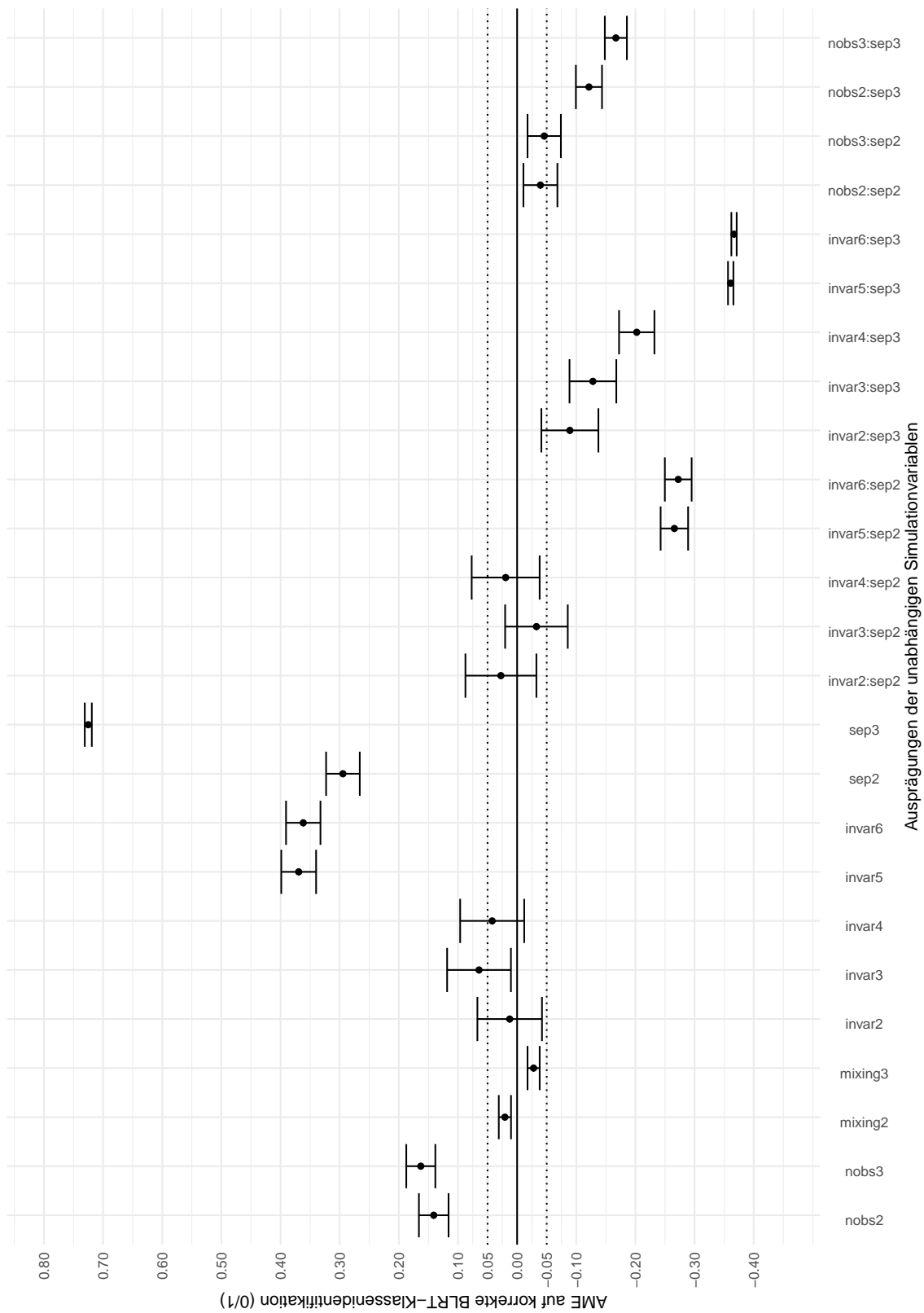


Abbildung 12: AMEs bzgl. der abhängigen Variablen der korrekten Klassenidentifikation nach BLRT.

Die Fallzahl der Schätzung beträgt 32399.

Prozentpunkte bessere absolute Klassifikationsrate im Vergleich zur kleinsten Distanz. Dies wird für die größte Klassendistanz von 4.0 MD noch einmal deutlich mit einer durchschnittlich um 73 Prozentpunkten besseren Klassifikationsrate im Vergleich zur Referenzkategorie übertroffen. Damit reagiert die Klassenidentifikation nach dem BLRT ähnlich stark auf die Klassendistanz wie der BIC.

Für die verschiedenen Abstufungen der Invarianzform der Messung zeigt sich, dass für geringe Klassendistanzen (0.8 MD) fehlende Messinvarianz im Durchschnitt sogar einen positiven Einfluss auf die absolute Rate der korrekten Klassenidentifikation hat, sofern die Faktorladungen von der fehlenden Invarianz betroffen (invar= 3, 5, 6) sind. Dies reduziert sich jedoch bereits deutlich für die mittlere Klassendistanzabstufung (1.5 MD) und ist für die höchsten Abstufungen (4.0 MD) nicht mehr vorhanden. Anhand der Interaktionsterme zwischen den Invarianzstufen und der Klassendistanz lässt sich erkennen, dass für mittlere Klassendistanzen (1.5 MD) die Invarianzstufen 5 und 6 mit gruppenspezifischer und kombinierter Nicht-Invarianz (bzgl. der manifesten Intercepts und Faktorladungen) die Klassenidentifikation negativ beeinflussen. Für die höchste Klassendistanz verstärkt sich diese Tendenz weiter und betrifft nicht nur die gruppenspezifischen Invarianz-Konfigurationen, sondern auch die längsschnittlichen Abweichungen von der Messinvarianz. Natürlich lässt sich die Interpretation der Interaktionsterme auch drehen, wobei der überaus positive Haupteffekt von hoher Klassendistanz (73 Prozentpunkte) für jede Abstufung der Messinvarianzen (egal ob längsschnittlich, gruppenspezifisch oder kombiniert) abgeschwächt wird. Wie sich bereits in der deskriptiven Auswertung unter ansonsten optimalen Bedingungen angedeutet hat, reagiert der BLRT nicht nur für die gruppenspezifischen und die kombinierten Invarianzabstufungen, sondern auch auf längsschnittliche Abweichungen. Der negative Effekt für die längsschnittlichen Abstufungen fällt aber deutlich geringer aus als für die gruppenspezifischen (und kombinierten) Konfigurationen (vgl. die Interaktionsterme in Abbildung 12).

9.5.3 Diskussion zur BLRT-Klassenidentifikation

Insgesamt zeigt die Anpassungsgüte des logistischen Metamodells zum BLRT mit einem *Nagelkerke* – $R^2 = 0.42$, dass die Klassenidentifikation, im Vergleich zu den IC-Maßen, moderat stark durch die Simulationsgrößen beeinflusst wird. Auch damit verortet sich der BLRT zwischen den beiden Maßen BIC und aBIC. Zugleich weist der BLRT unter optimalen Simulationsbedingungen ein ähnliches Grundniveau der Klassenidentifika-

tion auf wie der aBIC. Den größten Einfluss haben wiederum die Klassendistanz, die Invarianzstufen sowie die Interaktion zwischen beiden Größen. Dabei ist im Vergleich zu den IC-Maßen herauszustellen, dass für hohe Klassendistanzen nicht nur gruppenspezifische Non-Invarianzformen zu einer schlechteren Klassenidentifikation führen, sondern auch längsschnittliche Ausprägungen einen negativen Einfluss haben. Die Klassenanteilsverhältnisse zeigen im Mittel keinen relevanten verzerrenden Effekt und hohe Fallzahlen wirken sich lediglich positiv auf die Identifikation aus, wenn eine geringere Klassendistanz gegeben ist und führen bei hohen Klassendistanzen im Schnitt zur Identifikation von zu vielen Klassen. Diese Ergebnisse sind auch in Tabelle 15 im Überblick dargestellt.

Einschränkend muss erwähnt werden, dass die Performanz des BLRT evtl. dadurch eingeschränkt wird, dass zur Durchführung die Defaultoption von Mplus unter Verwendung eines sequentiellen Verfahrens angewendet wurde. Eine fest definierte hohe Ziehungszahl (z. B. 1000) hätte ggf. zu anderen/besseren Identifikationsraten geführt. Leider war dies aus zeitlichen bzw. rechentechnischen Beschränkungen im Rahmen dieser Untersuchung nicht anders möglich.⁵¹

⁵¹Durch die Verwendung des BLRT wird in dieser Studie sozusagen eine Resampling-Simulation innerhalb der Hauptsimulation durchgeführt, was die Rechenzeit exorbitant ansteigen lässt und dadurch ab einer gewissen Ziehungszahl auf dem BwUniCluster nicht mehr sinnvoll durchführbar war.

Tabelle 15: Zusammenfassung der BLRT-Performanz bzgl. der korrekten Klassenidentifikation

		BLRT
Grundniveau		+
Fehlende Messinvarianz	zeitlich	unbedeutend; für hohe Distanz: neg.
	gruppenspez.	für niedrige Distanz: pos.; für hohe Distanz: neg.
	kombiniert	für niedrige Distanz: pos.; für hohe Distanz: neg.
Hohe Klassendistanz		stark pos. beeinflusst
Hohe Fallzahl		für niedrige Distanz: pos.; für hohe Distanz: neg.
Kleine Klasse		unbedeutend
Einsatz-Empfehlung		Entscheidungskriterium bei suboptimalen oder unklaren Datensituationen (analog zu aBIC)

10 Ergebnisse zu Parameterschätzern

Die vorangegangenen Kapitel 8 und 9 stellten, neben Einzelheiten zur Simulationskonfiguration und zur konkreten Durchführung der Simulation, die Ergebnisse zur Klassensidentifikation anhand der relevanten IC-Indizes und des BLRT dar.

In den folgenden Kapiteln 10 und 11 sollen die Ergebnisse der Simulationsstudie hinsichtlich der tatsächlichen Wachstumsverläufe der latenten Klassen dargestellt und bewertet werden. Dabei werden zum einen die latenten Wachstumsparameter betrachtet, aber zum anderen auch die Schätzungen der Standardfehler in den Blick genommen, welche die Grundlage für Signifikanztests der Wachstumsverlaufsformen darstellen. Beruhend auf dem Einfluss der variierten Simulationsgrößen auf die Parameterschätzungen und Standardfehler sollen Handlungsempfehlungen für die Berechnung auf Basis von empirischen Daten abgeleitet werden.

Neben der korrekten Identifikation der tatsächlichen Anzahl der den Daten zugrundeliegenden latenten Klassen, müssen noch zusätzliche Leistungskriterien an MIGMMs gestellt werden, um sie für die empirische Forschungspraxis sinnvoll und zuverlässig einsetzen zu können. Wenn die Anzahl der distinkten Wachstumsverläufe feststeht, stellt sich als Nächstes die Frage, ob die identifizierten Verläufe auch in ihrer Form korrekt rekonstruiert werden können. Die Verläufe der MIGMMs lassen sich über die Ko-/Varianzen und die Mittelwerte der latenten Wachstumsparameter charakterisieren. Daher werden Verzerrungen in diesen Größen auch im Weiteren als Performanzkriterien herangezogen. Für den «Bias» bzw. den skalenunabhängigen «relative Bias» werden die in der Simulationsliteratur üblichen Grenzwerte von $< 5\%$ (vgl. Hoogland & Boomsma, 1998: 346) sowie die weichere Grenze von $< 10\%$ ⁵², welche in GMM-Simulationen bereits Anwendungen gefunden hat (vgl. z. B. Kim & Kim, 2012: 466; Depaoli, 2013: 197), angesetzt. Für die Berechnung der Verzerrungen wurden diejenigen Replikationen herangezogen, welche von mindestens einem der IC-Indizes korrekt als 3-Klassenmodell identifiziert werden konnten. Dies lässt sich dadurch rechtfertigen, dass Verzerrungen der Parameter für MIGM-Modelle erst dann substantiell relevant werden, wenn die Anzahl der Klassen korrekt identifiziert werden kann. Ansonsten würden die Wachstumsverläufe einer 3-Klassenlösung gar nicht erst betrachtet werden, da die Wachstumsverläufe der besser an die Daten angepassten Klassenmodelle zu interpretieren wäre. Somit sollten die weiter untersuchten Replikatio-

⁵²Da Verzerrungen sowohl negativ als auch positiv ausfallen können, muss strenggenommen von $\pm 5\%$ und $\pm 10\%$ gesprochen werden.

nen zumindest eine minimale Chance haben im Rahmen der Modellschätzung als korrekte Klassenlösung identifiziert zu werden, was im Weiteren durch den Zuspruch von mindestens einem der IC-Indizes⁵³ operationalisiert wird. Dies führt in der Summe zu einer Reduzierung der betrachteten Anzahl der Replikationen der Studie von 48594 auf 23084. Daher werden in den folgenden tabellarischen Aufschlüsselungen auch die zugrundeliegenden Fallzahlen der einzelnen Konfigurationen ausgewiesen.

10.1 Haupteffekte bzgl. des Bias

Wie bereits für die Auswertungen zur Klassenidentifikation werden zunächst die Haupteffekte der einzelnen Simulationsgrößen unter Fixierung der restlichen Variablen dargestellt. Die konstant gehaltenen Größen werden dabei wiederum auf Ausprägungen der Simulationsvariablen fixiert, welche die fehlerfreie Berechnung der Wachstumsparameter begünstigen sollten. Für die folgenden Darstellungen zum relativen Bias werden – sofern dies nicht die jeweils analysierte Variable darstellt – die Invarianzstufe auf völlige skalare Invarianz der Messung (Invar=1), die Fallzahl auf 1300 Fälle (Nobs=3), die Klassenanteilsverhältnisse auf gänzlich Gleichverteilung über die Klassen (Mixing=1) und die Klassendistanz auf 4.0 MD festgelegt (Sep=3).⁵⁴ Der relative Bias wird jeweils für die Varianz- und Kovarianzen aber auch für die Means der Intercepts und Slopes der latenten Klassen dargestellt.⁵⁵

Invarianzstufen. Für die verschiedenen Abstufungen der Messinvarianz zeigt sich hinsichtlich der Varianzparameter der Wachstumsfaktoren, dass fehlende Messinvarianz hinsichtlich der Intercepts alleine noch nicht zu inakzeptablen Verzerrungen führt. Betrachtet man die relevanten Kategorien der Invarianz (2, 4) in Tabelle 16 zeigt sich, dass die Werte des relativen Bias unter der konservativen Schwelle von 5 % bleiben.⁵⁶ Die Varianzschätzungen sind folglich, unter ansonsten optimalen Simulationsbedingungen, sowohl gegenüber längsschnittlicher wie auch gegenüber gruppenspezifischer Invarianz der Intercepts robust.

⁵³Die Identifikation nach dem BLRT kann hier nicht berücksichtigt werden, da der Test nicht in äquivalenter Weise für alle 300 Replikationen innerhalb einer Konfiguration durchgeführt wurde.

⁵⁴Referenztabellen, welche alle Simulationskonfigurationen umfassen, können im Anhang in den Tabellen 47 und 48 gefunden werden.

⁵⁵Lediglich der Mean des Intercepts der Klasse 3 wird nicht dargestellt, da sich durch den Populationswert von 0 der relative Bias nicht mathematisch sinnvoll berechnen lässt.

⁵⁶Ausnahme ist die Kovarianz in Invarianzkategorie 2, welche die konservative Grenze verletzt, aber immer noch unter der liberaleren Grenze von 10 % bleibt.

Dieses Bild wandelt sich allerdings, wenn nicht nur die Intercepts fehlende Invarianz aufweisen, sondern zusätzlich die Faktorladungen der manifesten Indikatoren als nicht invariant spezifiziert wurden. Alle Ko-/Varianzparameterschätzungen übersteigen für diese Konfigurationen (3, 5, 6) die konservative Grenze. Mit lediglich einer Ausnahme wird auch die weiche Grenze von 10 % relativem Bias überschritten. Besonders deutlich fallen die Verzerrungen der Ko-/Varianzparameter aus, wenn längsschnittliche und gruppenspezifische Non-Invarianz kombiniert werden. Im Vergleich zu den beiden einzelnen Invarianzausprägungen steigt der Grad der Verzerrung der Ko-/Varianzparameter in der Kombination beider Messinvarianztypen um den Faktor 2 bis 4.

Tabelle 16: Relativer Bias nach Invarianz-Stufen

Invar	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(S_C1)	M(S_C2)	M(S_C3)	N
1.000	0.000	0.004	-0.002	-0.002	-0.004	0.000	0.000	0.006	300
2.000	0.037	0.032	0.059	0.048	0.076	-0.009	-0.035	-0.463	300
3.000	0.149	-0.079	0.130	0.138	0.130	-0.036	-0.058	-0.432	300
4.000	0.043	0.038	-0.029	-0.268	0.323	-0.456	0.915	0.010	240
5.000	0.106	0.120	0.148	0.494	0.067	-0.359	0.441	0.193	152
6.000	0.262	0.286	0.621	0.717	0.143	-0.299	0.282	0.225	144

Info: Klassendistanz=4 MD, Mixing=33.33.33, Nobs=1300.

Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

Für den Bias der Mittelwerte der Wachstumsfaktoren muss zwischen den Intercept- und Slopefaktoren der einzelnen latenten Klassen unterschieden werden. Da sich das Muster der Fehlspezifikationen zwischen den einzelnen latenten Klassen unterscheidet (vgl. hierzu die Populationswerte in Kapitel 8), müssen die Slope- und Interceptfaktoren auch hiernach getrennt aufgeschlüsselt werden. Die beiden hier betrachteten Intercept-Mittelwerte der Klassen 1 und 2 zeigen⁵⁷ für eine längsschnittliche negativ gerichtete Non-Invarianz der Intercepts der manifesten Indikatoren positiv verzerrte Schätzwerte. Diese Verzerrungen bewegen sich nahe oder überschreiten die konservative Verzerrungsgrenze von 5 %; bleiben aber klar unter der weicheren Grenze von 10 %. Erst in der zweiten längsschnittlichen Nicht-Invarianzkonfiguration, welche sowohl die Intercepts der manifesten Variablen als auch die Faktorladungen negativ nicht-invariant setzt, wird die weiche Grenze von 10 % durch eine positive Verzerrung überschritten. Folglich addiert sich fehlende längsschnittliche Invarianz der Intercepts und der Faktorladungen auf und führt, zumindest in der Kombination, zu nicht mehr akzeptablen Verzerrungen der Interceptfaktoren der Wachstumsmodelle.

⁵⁷Der Interceptfaktor von Klasse 3 wird aufgrund bereits genannter Gründe nicht betrachtet, vgl. Fußnote 55.

Auch für die gruppenspezifische Non-Invarianz zeigen sich Verzerrungen der Interceptfaktoren. Diese Verzerrungen fallen deutlich stärker aus als in den längsschnittlichen Konfigurationen.⁵⁸ Variante Intercepts zwischen den Klassen führen zu deutlich verzerrten Schätzungen der Interceptfaktoren, welche sowohl die konservative als auch die weiche Grenze für den relativen Bias mit Werten um 27 und 32 % deutlich überschreiten. Diese Verzerrungen entsprechen in den Vorzeichen der Richtung der spezifizierten Invarianz. Positive Abweichungen der Intercepts führen zu positiven Verzerrungen und negative Abweichungen in den latenten Klassen führen zu negativen Verzerrungen. Wenn nun zusätzlich die Faktorladungen innerhalb des Populationsmodells der latenten Klassen als nicht-invariant spezifiziert werden, zeigt sich, dass die durchschnittliche Schätzung der Interceptfaktoren nicht etwa wie bei den längsschnittlichen Konfigurationen verstärkt, sondern (über-)kompensiert wird. So bleibt die durchschnittliche Schätzung des Interceptfaktors der ersten Klasse zwar im nicht akzeptablen Bereich, aber die Richtung der Verzerrungen dreht sich von negativ zu positiv. Der äquivalente (aber bzgl. der Vorzeichen umgedrehte) Zusammenhang zeigt sich in Klasse 2. Die zuvor positiv gerichtete Verzerrung des Interceptfaktors wird durch die zusätzliche nicht-invariante Spezifikation der Faktorladungen kompensiert und bewegt sich damit im noch akzeptablen Bereich (nach der weichen Grenze). Gruppenspezifische Nicht-Varianzen bzgl. der manifesten Intercepts und der Faktorladungen können sich folglich kompensieren sofern sie bzgl. der Vorzeichen gleichgerichtet sind.⁵⁹

In der Kombination von längsschnittlicher und gruppenspezifischer Nicht-Invarianz zeigt sich, dass für beide betrachteten Interceptfaktoren das Ausmaß der Verzerrung zunimmt und somit nicht von einer wesentlichen Kompensationsleistung durch die Kombinationen beider Nicht-Invarianzformen auszugehen ist.

Neben den Interceptfaktoren wird die Wachstumsform der latenten Verlaufsklassen zusätzlich durch die Slopefaktoren bestimmt. Für die längsschnittliche Nicht-Invarianz der Intercepts zeigt sich keine bedeutende Verzerrung des Slopefaktors für Klasse 1 und Klasse 2. Allerdings ist der Slopefaktor von Klasse 3 deutlich um 46 % negativ verzerrt. Das gleiche Muster zeigt sich, wenn zusätzlich die längsschnittlichen Faktorladungen im Populations-

⁵⁸Dabei gilt es zu beachten, dass der Umfang der Abweichungen der Faktorladungen und Intercepts in den gruppenspezifischen Konfigurationen in etwa dem Unterschied zwischen dem ersten und dem letzten Erhebungszeitpunkt der längsschnittlichen Konfigurationen entspricht.

⁵⁹Dass dieser Zusammenhang auch umgekehrt, sprich eine Verstärkung der Verzerrung durch entgegengesetzte Vorzeichen der Intercept- und Faktorladungabweichungen, gilt, scheint zwar plausibel, liegt aber aufgrund der Simulationskonfiguration prinzipiell außerhalb der im Rahmen dieser MC-Studie ableitbaren Schlussfolgerungen.

modell nicht-invariant spezifiziert werden. Dabei nähern sich die Verzerrungen der Slopes in Klasse 1 und Klasse 2 der konservativen Grenze von 5 % an oder überschreiten diese leicht. Der durchschnittliche Slope für Klasse 3 bleibt beinahe unverändert stark negativ verzerrt und damit auch deutlich außerhalb des akzeptablen Intervalls der weichen Grenze von 10 %. Die Richtung der Verzerrung entspricht dabei dem Vorzeichen der Abweichung im Rahmen der Nicht-Invarianz, was analoge Simulationsergebnisse von Zhang (2016) repliziert (vgl. Zhang, 2016: 76).

Für die gruppenspezifischen Konfigurationen dreht sich dieses Bild zunächst völlig. In der Invarianzkonfiguration 4 sind die durchschnittlichen Abweichungen für die Klassen 1 und 2 mit knapp 46 und 92 % deutlich außerhalb des akzeptablen Bereichs, wohingegen die Referenzklasse 3 mit einem Prozent sowohl nach der weichen als auch nach der konservativen Grenze unverzerrt ist. Äquivalent zur Richtung der Abweichungen in den Intercepts ist der Slopefaktor für Klasse 1 negativ und der Slopefaktor für Klasse 2 positiv verzerrt. Werden nun zusätzlich die Faktorladungen in Invarianzstufe 5 nicht-invariant generiert, nimmt der Grad der Verzerrung der Slopefaktoren in den Klassen 1 und 2 wieder ab, ohne dass dadurch allerdings akzeptable Verzerrungsraten erreicht werden. Diese Abnahme in den Slopefaktoren für die ersten beiden Klassen wird allerdings damit erkauft, dass der durchschnittliche Slopeparameter der unveränderten⁶⁰ Referenzklasse 3 mit knapp 20 % relativen Bias positiv verzerrt geschätzt wird. Auch für die sechste Invarianzkategorie, welche die längsschnittliche und gruppenspezifische Nicht-Invarianz kombiniert, zeigt sich eine nochmalige Reduktion der Verzerrung der Slopefaktoren für Klasse 1 und 2, wobei allerdings die Verzerrung für Klasse 3 gleichzeitig zunimmt. Insgesamt ist damit aber für die 6. Simulationkategorie keine der durchschnittlichen Slopeparameterschätzungen im noch akzeptablen Bereich des Intervalls zwischen ± 10 %.

Bei all diesen Ergebnissen muss einschränkend beachtet werden, dass insbesondere den gruppenspezifischen Konfigurationen durch die geringere Anzahl an einbezogenen Replikationen (auf Grundlage der korrekten Klassenidentifikation) ein höheres Ausmaß an Unsicherheit zukommt.

Alles in allem kann zusammenfassend konstatiert werden, dass nicht-invariante Faktorladungen in Kombination mit nicht-invarianten Intercepts für nahezu alle Parameterklassen der Wachstumsform ein Problem hinsichtlich des Bias darstellen. Dabei ist es unerheblich, ob es sich um längsschnittliche oder gruppenspezifische Non-Invarianz handelt. Auch die

⁶⁰ «Unverändert» nach den Festlegungen der Populationsmodelle.

Kombination aus längsschnittlicher und gruppenspezifischer Nicht-Invarianz führt eher zu einer Verstärkung der Parameterverzerrungen, so dass die Annahme von evtl. Kompensationsleistungen bei entgegengesetzter Richtung von längsschnittlicher und gruppenspezifischer Nicht-Invarianz insgesamt nicht haltbar bleibt. Lediglich die Schätzung der Varianzparameter scheint gegenüber der Nicht-Invarianz der manifesten Intercepts robust zu sein. Diese Aussagen gelten für ansonsten optimale Simulationsbedingungen, welche eine unverzerrte Berechnung der Wachstumsparameter begünstigen sollten und unter der Voraussetzung der vorangehenden korrekten Klassenidentifikation. In Übereinstimmung mit gruppenspezifischen Analysen nach Kim & Wang (2017) zeigt sich, dass die Verzerrungen der Wachstumsform deutlich mit den Nicht-Invarianzformen assoziiert sind (vgl. Kim & Wang, 2017: 15).⁶¹ Selbst wenn die korrekte Klassenanzahl unter verschiedenen Invarianzformen extrahiert wird, kann folglich nicht davon ausgegangen werden, dass die Form des Wachstumsverlaufs adäquat rekonstruiert werden kann.

Klassendistanz. Wie bereits bei der Analyse der Klassenidentifikation zeigt sich für die Klassendistanz eine starke Einflussbeziehung. Für die durch die IC-Indizes korrekt als 3-Klassen-Modelle identifizierten Replikationen zeigen sich bei der Klassendistanz von 4.0 MD (sep=3) keinerlei nennenswerte Verzerrungen in den Ko-/Varianz- und den Mittelwertparametern der Wachstumsfaktoren. Der relative Bias überschreitet in dieser Konfiguration nicht einmal die 1-%-Marke und bleibt somit deutlich unter der konservativen Grenze von 5 % (vgl. Tabelle 17). Für die mittlere Klassendistanz von 1.5 MD zeigt sich für die Varianzwerte der Wachstumsfaktoren auch noch keine nennenswerte Verzerrung, da die harte Grenze des relativen Bias noch nicht überschritten wird. Für die Kovarianz gilt allerdings nicht das Gleiche. Für die mittlere Klassendistanz wird mit einem relativen Fehler von ca. 52 % sowohl die harte wie auch die weiche Grenze für Parameterverzerrungen deutlich überschritten. Da die Kovarianz von Wachstumsparametern in der empirischen Praxis allerdings selten von inhaltlichem Interesse sein sollte, stellt dies ggf. kein essenzielles Problem für die Performanz von MIGMMs dar.

Anders verhält es sich mit den Mittelwerten der Intercept- und Slopefaktoren der la-

⁶¹Einschränkend muss allerdings erwähnt werden, dass in dieser Studie keine klare Korrespondenz zwischen den nicht-invarianten Intercepts und Interceptfaktorverzerrungen sowie zwischen Faktorladung und Slopefaktoren (vgl. Kim & Wang, 2017: 9 f.) gefunden werden konnte, wie bei Kim & Wang (2017). Diese Differenzen könnte im unterschiedlichen Aggregationsgrad der Auswertung begründet sein, könnte aber auch mit der anders spezifizierten Populationsschätzmodellen zusammenhängen, da z. B. in dieser Studie, im Gegensatz zu Kim & Wang (2017), Fehlerkovarianzen zugelassen wurden.

tenten Klassen. Da diese Werte die durchschnittlichen Start- und Entwicklungsverläufe wiedergeben, sind sie für empirische Auswertung höchst relevant und in aller Regel von substanziellem inhaltlichen Interesse. Allerdings zeigt sich bereits für die mittlere Klassendistanz, dass die Mittelwerte der Intercept- und die Slopefaktoren der einzelnen Klassen deutlich verzerrt geschätzt werden. Dabei wird sowohl tlw. die harte wie auch die weiche Grenze für den relativen Bias überschritten und es zeigt sich mit Abweichungen von 6 bis 103 % eine große Spannweite.

Tabelle 17: Relativer Bias nach Klassendistanz-Stufen

Sep	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(S_C1)	M(S_C2)	M(S_C3)	N
1.000	-0.231	-0.135	0.153	1.356	0.005	-0.057	0.090	0.566	62
2.000	-0.028	-0.014	0.523	0.439	0.060	-0.087	0.171	1.039	91
3.000	0.000	0.004	-0.002	-0.002	-0.004	0.000	0.000	0.006	300

Info: Invarianz=1, Mixing=33.33.33, Nobs=1300.

Var()=Varianz; KoVar()=Kovarianz; M(Y_CX)=Mean des Wachstumsparameters Y der Klasse X

Für die geringste Klassendistanz mit 0.8 MD (sep=1) sind dann schließlich auch die Varianzparameter der Intercept- und Slopefaktoren deutlich nach unten verzerrt und überschreiten auch die weiche Grenze von 10%. Dabei gilt es insbesondere bei der Simulationskonfiguration mit geringer Klassendistanz zu berücksichtigen, dass die Schätzergebnisse zur relativen Verzerrung mit einer größeren Unsicherheit behaftet sind. Da die Vorstufe der korrekten Klassenidentifizierung die Mehrzahl der Replikationen ausschließt, basiert die Berechnung des relativen Bias hier lediglich auf 62 Replikationen, wodurch der Einfluss der Zufallsziehung stärker zum Tragen kommt wie in den anderen Simulationskonfigurationen.

Insgesamt zeigt dies, dass selbst bei korrekter Identifikation der Anzahl der latenten Klassen die Verlaufsmuster selbiger Entwicklungsklassen bei einer geringen Klassendistanz verzerrt geschätzt werden. Bereits bei einer mittleren Klassendistanz von 1.5 MD kann dies zu verzerrten Schlussfolgerungen bzgl. der durchschnittlichen Wachstumsverlaufsform führen. Bei als klein einzustufenden Klassendistanzen wird zusätzlich auch die Heterogenität dieser Verlaufsformen im Schnitt verzerrt eingeschätzt werden. Diese Ergebnisse deuten darauf hin, dass Forscher/Forscherinnen bei der Interpretation von Parameterschätzungen vorsichtig sein sollten, wenn der Abstand zwischen den Klassen gering ist. Insgesamt entspricht dies den Ergebnissen von Koo & Leite (2014) für einfache GMMs (vgl. Koo & Leite, 2014: 215). Diese allgemeine Tendenz im Zusammenhang zwischen Klassendistanz und Bias der Parameter können auch bei Serang et al. (2016) gefunden werden,

wobei der Grad der Verzerrungen aufgrund der insgesamt recht hohen Klassendistanzen der Simulation weniger prägnant ausfällt (vgl. Serang et al., 2016: 6 ff.). Der Einfluss der Klassendistanz bzgl. der Parameterschätzungen der Wachstumsparameter kann somit, anhand der hier dargestellten Ergebnisse, auf SOGMMs vorläufig ausgeweitet werden.

Fallzahl und Klassenanteilsverhältnisse. Wie aus den Tabellen 18 und 19 entnommen werden kann, haben die Fallzahl und die Verteilung der Fälle über die Klassen, unter ansonsten optimalen Bedingungen, keinen nennenswerten negativen Einfluss auf die Varianz- und Mittelwertparameter der Wachstumsfaktoren.

Tabelle 18: Relativer Bias nach Fallzahl-Stufen

Nobs	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(S_C1)	M(S_C2)	M(S_C3)	N
1.000	-0.014	-0.029	-0.030	0.004	-0.006	0.001	0.002	0.023	297
2.000	-0.002	-0.000	0.008	0.001	-0.004	0.001	0.002	0.014	299
3.000	0.000	0.004	-0.002	-0.002	-0.004	0.000	0.000	0.006	300

Info: Invarianz=1, Mixing=33.33.33, Klassendistanz=4 MD.

Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

Die kritischen Schwellenwerte von 5 oder 10 % relativer Verzerrung werden für keine untersuchten Parameter erreicht. Für die Fallzahlausprägung von 300 Fällen (nobs=1) lässt sich anhand von Tabelle 18 erahnen, dass sich die Varianz- und Kovarianzschätzungen bei geringeren Fallzahlen der harten Grenze von 5 % annähern würden. Allerdings ist dies eine prinzipiell nicht zulässige Extrapolation über die tatsächliche Aussagekraft dieser MC-Studie hinaus. Der fehlende Einfluss unter ansonsten günstigen Simulationsbedingungen fügt sich in den bisherigen Forschungsstand zu GMMs ein und kann somit auch auf MIGMMs ausgeweitet werden. So konnte auch Kim & Kim (2012) für klassische GMMs unter ähnlichen Simulationsbedingungen einen kritischen Mindestfallzahlwert von 200 ausmachen, wobei allerdings erwähnt werden muss, dass dieser Mindestwert unter ungünstigeren Simulationsbedingen sehr schnell ansteigt (vgl. Kim & Kim, 2012: 466 f.). Für die Klassenanteilsverhältnisse lässt sich hingegen keinerlei Tendenz entdecken. Diese Ergebnisse werden auch nicht durch über die Klassenidentifikation ausgeschlossene Replikationsfälle beeinflusst, da alle betrachteten Simulationskonfigurationen >297 Replikationen beinhalten und damit nahe oder genau auf dem Maximum pro Simulationskonfiguration liegen. Insgesamt muss konstatiert werden, dass weder Fallzahl noch die Anteilsverhältnisse unter ansonsten günstigen Datenbedingungen einen eigenen verzerrenden Einfluss auf die Varianz- und Parameterschätzungen von MIGMMs ausüben. Sofern also die korrekte Klassenanzahl identifiziert werden konnte, üben die zugrundeliegende

Tabelle 19: Relativer Bias nach Mixing-Stufen

Mixing	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(S_C1)	M(S_C2)	M(S_C3)	N
1.000	0.000	0.004	-0.002	-0.002	-0.004	0.000	0.000	0.006	300
2.000	-0.000	0.002	-0.002	-0.002	-0.003	0.000	-0.001	0.007	300
3.000	-0.001	-0.003	-0.009	-0.001	-0.011	0.000	-0.006	0.003	299

Info: Invarianz=1, Nobs=1300, Klassendistanz=4 MD.

Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

Fallzahl sowie die Verteilung dieser Fälle über die Klassen keinen eigenen nennenswerten verzerrenden Effekt auf die einzelnen Klassenverlaufsmuster aus.

10.2 Modellbasierte Analyse bzgl. des Bias

Für die modellbasierte Auswertung hinsichtlich der Verzerrungen der Wachstumsparameter wird der rohe Bias (auch: Raw Bias) der jeweiligen Parameter als abhängige Größe berücksichtigt. Um die Ergebnisinterpretation zu vereinfachen, wird dabei der absolute Betrag des Bias als abhängige Variable berücksichtigt. Negative Schätzwerte der Steigungskoeffizienten markieren damit geringere Verzerrungen, wohingegen positive Steigungskoeffizienten für ein höheres Maß an Parameterverzerrung stehen. In die Berechnungen gehen wiederum nur die Replikationen ein, welche mindestens von einem der IC-Indizes korrekt als 3-Klassenlösung identifiziert wurden. Für jeden betrachteten Wachstumsparameter wird hier ein eigenes lineares Modell geschätzt, welches jeweils die Haupt- und Interaktionseffekte der Simulationsgrößen beinhaltet. Für die Berechnung der Signifikanztests wurden «Eicker-Huber-White robust standard errors» aus der Familie der Sandwich-Estimators⁶² verwendet, da sich bei einigen der Metamodelle in der Modell-Diagnostik Probleme hinsichtlich der Heteroskedastizität gezeigt haben.⁶³ Für die Interaktionsterme werden nur Effekte berücksichtigt, welche mindestens 2 % (ΔR^2) mehr Varianzaufklärung zu den Haupteffekten beitragen, was nach Cohen (1988) einer kleinen Effektstärke entspricht (vgl. Cohen, 1988: 413 ff.). Die Signifikanz einzelner Dummy-Effekt-Terme wurde dabei nicht als alleinige Leitschnur herangezogen, da die Fallzahl mit 23 084 Fällen immer noch sehr hoch ausfällt, womit auch substantiell unbedeutende Effekte als signifikant angezeigt werden. Die Ergebnisse der folgenden Metamodelle sind mit unstandardisierten

⁶²Für die Berechnung der Standardfehler wurde die Variante «HC3» nach Davidson & MacKinnon (1993) gewählt, welche in Simulationsstudien vergleichsweise gute Ergebnisse lieferte (vgl. Hayes & Cai, 2007: 713).

⁶³Die visuelle Diagnostik zu allen Meta-Modellen kann aus dem Online-Appendix A.5 entnommen werden. Trotz der Bildung von Interaktionstermen wurden die relevanten Grenzwerte für Multikollinearität für keines der hier präsentierten Modelle überschritten.

Koeffizienten in den Tabelle 20, 22 und 25 dargestellt.⁶⁴ Die relative Bedeutsamkeit der Effekte ist in den Tabellen 21, 23 und 24 mit den Delta- R^2 -Werten ausgewiesen.

10.2.1 Analyse bzgl. des Bias Ko-/Varianzparameter

Die Varianzparameter der Wachstumsfaktoren beschreiben den Grad an Unterschiedlichkeit, welcher hinsichtlich der Verlaufsmuster innerhalb der latenten Klassen besteht. Verzerrungen der Varianzschätzungen würden somit mit einer Über- oder einer Unterschätzung der tatsächlichen Heterogenität innerhalb der Klassen einhergehen.

Während bei der deskriptiven Auswertung der Haupteffekte die Fallzahl und die Klassenanteilsverhältnisse keinen nennenswerten Einfluss auf die Varianz und Kovarianzparameter aufwiesen, zeigt die modellbasierte Betrachtung über alle Simulationskonfigurationen ein etwas differenzierteres Bild. Die Klassenanteilsverhältnisse zeigen wie zuvor (unter optimalen Simulationsbedingungen) keinen nennenswerten Einfluss auf die Schätzungen der Ko-/Varianzen. Mit einem Delta- R^2 von maximal 0.2 % (vgl. Tabelle 21 auf Seite 133) wird keine der Varianzschätzungen substantiell beeinflusst, sofern die korrekte Klassenanzahl identifiziert wurde. Für die Fallzahl weicht die modellbasierte Auswertung jedoch vom Muster der deskriptiven Tabellenanalyse ab. Mit einem Delta- R^2 zwischen 4 und 5 % kann die Fallzahl die Varianz der Ko-/Varianzschätzungen zusätzlich substantiell statistisch erklären. Je höher die Fallzahl ausfällt, desto geringer sind die Verzerrungen in den Varianzschätzungen in allen drei Metamodellen (vgl. Tabelle 20 auf Seite 132).

Für die Klassendistanz zeigt sich sowohl in Richtung als auch bzgl. der Stärke ein ähnlich gearteter Effekt hinsichtlich der Ko-/Varianzschätzungen. Mit einer zusätzlichen Varianzaufklärung von ca. 4 bis 4.5 % bewirkt eine höhere Klassendistanz geringere Verzerrungen für alle Varianzschätzungen.

Für die Haupteffekte und die Interaktionen der Invarianzstufen zeigt sich ein weniger einheitliches Bild über die Metamodelle hinweg. Daher werden im Folgenden die Effekte für die einzelnen Modelle und Schätzparameter getrennt dargestellt:

- Für die Varianzschätzung des Interceptfaktors (vgl. Tabelle 20: Modell 1 auf Seite 132) hat der Haupteffekt der Invarianzstufen, unter Kontrolle aller restlichen Einflussgrößen, keinen substantiell eigenen Einfluss, mit einer zusätzlichen Varianzaufklärung um lediglich 0.4 %. In der Tendenz zeigt der Haupteffekt für Invarianzstufen

⁶⁴Die Ergebnistabellen wurden mit dem Softwarepaket «stargazer» erstellt: Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. <https://CRAN.R-project.org/package=stargazer>.

Tabelle 20: Regressions-Ergebnisse zu Bias der Ko-/Varianz Parameter (robuste SE; HC3)

	<i>Dependent variable:</i>		
	model_bias_i_v	model_bias_s_v	model_bias_is_cv
	Pop.Val.: 0.25	Pop.Val.: 0.04	Pop.Val.: 0.02
	(1)	(2)	(3)
Constant	0.092*** (0.003)	0.019*** (0.001)	0.027*** (0.001)
nobs2	-0.013*** (0.001)	-0.003*** (0.0001)	-0.004*** (0.0002)
nobs3	-0.019*** (0.001)	-0.005*** (0.0001)	-0.007*** (0.0002)
mixing2	-0.003*** (0.001)	-0.0001 (0.0001)	-0.001*** (0.0002)
mixing3	-0.005*** (0.001)	-0.0001 (0.0001)	-0.001*** (0.0002)
invar2	-0.011** (0.004)	0.001 (0.001)	0.0003 (0.001)
invar3	-0.011** (0.003)	0.004*** (0.001)	-0.002 (0.001)
invar4	-0.002 (0.004)	-0.0001 (0.001)	-0.0005 (0.001)
invar5	-0.001 (0.003)	-0.005*** (0.001)	-0.008*** (0.001)
invar6	-0.013*** (0.003)	0.004*** (0.001)	-0.008*** (0.001)
sep2	-0.039*** (0.003)	-0.003*** (0.001)	-0.005*** (0.001)
sep3	-0.061*** (0.003)	-0.012*** (0.0005)	-0.017*** (0.001)
invar2:sep2	0.007 (0.004)	-0.002** (0.001)	-0.001 (0.001)
invar3:sep2	0.023*** (0.004)	-0.003** (0.001)	0.004** (0.001)
invar4:sep2	-0.006 (0.004)	-0.002** (0.001)	-0.004*** (0.001)
invar5:sep2	0.009* (0.004)	0.005*** (0.001)	0.011*** (0.001)
invar6:sep2	0.023*** (0.004)	-0.003*** (0.001)	0.012*** (0.001)
invar2:sep3	0.013*** (0.004)	-0.0005 (0.001)	-0.0001 (0.001)
invar3:sep3	0.030*** (0.003)	-0.002** (0.001)	0.003** (0.001)
invar4:sep3	0.007 (0.004)	0.001 (0.001)	0.001 (0.001)
invar5:sep3	0.034*** (0.003)	0.009*** (0.001)	0.013*** (0.001)
invar6:sep3	0.082*** (0.003)	0.007*** (0.001)	0.020*** (0.001)
Observations	23,084	23,084	23,084
R ²	0.314	0.345	0.272
Adjusted R ²	0.313	0.345	0.271
Residual Std. Error (df = 23062)	0.034	0.007	0.011
F Statistic (df = 21; 23062)	502.744***	579.323***	410.575***

Info:

*p<0.05; **p<0.01; ***p<0.001

Tabelle 21: Delta-R-Quadrat der Praediktoren zum Bias der Ko-/Varianz Parameter

	df	dR Var I	dR Var S	dR KoVar IS
nobs	2	0.037	0.045	0.049
mixing	2	0.002	0.00002	0.001
invar	5	0.004	0.030	0.016
sep	2	0.044	0.042	0.040
invar:sep	10	0.065	0.042	0.053

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 20.

mit längsschnittlichen Verzerrungen sogar eher eine Reduzierung der Verzerrung. Diese Tendenz wandelt sich jedoch unter Mitbetrachtung des Interaktionseffekts zwischen den Invarianzstufen und der Klassendistanz. Für höhere Klassendistanzen nimmt der Grad der Verzerrung durch die Invarianzstufen deutlich zu. Insbesondere für Konfigurationen, welche nicht-invariante Faktorladungen enthalten, erhöht sich der Grad des Varianzbias des Interceptfaktors. Die zusätzliche Varianzaufklärung durch den Interaktionseffekt beträgt dabei 6.5 %. Nach dem Metamodell beträgt der geschätzte rohe Bias für die kombinierte Invarianzstufe 6 bei maximaler Klassendistanz insgesamt 0.069, was in diesem Fall einem relativen Bias von 27.6 % entspricht und damit auch außerhalb der weichen Grenzwerte liegt.

- Für die Schätzung der Varianz des Slopefaktors (vgl. Tabelle 20: Modell 2 auf Seite 132) hat auch der Haupteffekt der Invarianzstufen mit einem Delta- R^2 von 3 % einen interpretationswürdigen Einfluss. Simulationkonfigurationen mit nicht-invariant spezifizierten Faktorladungen haben dabei einen Einfluss auf die Varianz des Slopefaktors. Invarianzformen, welche lediglich hinsichtlich der Intercepts der manifesten Variablen nicht-invariant spezifiziert wurden, beeinflussen die Schätzung der Varianz des Slopeparameters hingegen eher nicht. Für Faktorladungs-Konfigurationen mit längsschnittlicher Nicht-Invarianz nehmen die Verzerrungen zu, wohingegen die Varianz des Slopes für gruppenspezifische Nicht-Invarianzen sogar weniger verzerrt geschätzt werden. Unter Berücksichtigung des Interaktionseffekts zwischen den Invarianzformen und der Klassendistanz, was auch zu einem statistischen Erklärungszuwachs um 4.2 % führt, werden weitere Muster ersichtlich. So zeigt sich, dass unter einer mittleren Klassendistanz die Über- und Unterschätzung der Varianz des Slopeparameters die Waage halten. Infolgedessen zeigt sich bei einer mittleren Klassendistanz von 1.5 MD weder ein positiver noch ein negativer Effekt auf die Varianzschätzung; und dies unabhängig von der Form der Nicht-Invarianz. Unter großer

Klassendistanz von 4.0 MD wandelt sich dieses Bild wieder etwas. Längsschnittliche Invarianzformen haben unter großer Klassendistanz wiederum keinen Einfluss auf die Varianz des Slopefaktors. Anders verhält es sich allerdings bei der gruppenspezifischen und kombinierten Invarianzkonfigurationen, welche die Faktorladungen beinhalten. Diese führen unter großer Klassendistanz zu einer größeren Verzerrung des Slopefaktors.

- Für die Kovarianz zwischen dem Intercept- und dem Slopefaktor (vgl. Tabelle 20: Modell 3 auf Seite 132) zeigt sich ein geringfügiger Haupteffekt der Invarianzform (Delta- R^2 : 1.6 %). Für die Invarianzstufen 5 und 6 wird die Kovarianz sogar weniger verzerrt geschätzt. Aber auch dieser Effekt dreht sich für höhere Klassendistanzen, wie durch die zusätzliche Betrachtung des Interaktionseffekts deutlich wird (Delta- R^2 : 5.3 %). Ab einer Distanz von 4.0 MD dreht sich der Gesamteffekt und zeigt z. B. für Invarianzstufe 6 eine geschätzte Gesamtverzerrung von 0.012, was bei einem Populationswert von 0.02 einem relativen Fehler von 60 % entspricht.

Alles in allem zeigt sich für die Invarianzstufen und die Interaktion mit der Klassen-separation, dass für hohe Klassendistanzen der Grad der Verzerrung der Ko-/Varianzen durch die Invarianz insbesondere in den Invarianzstufen 5 und 6 allgemein zunimmt. Für die Invarianzstufe 2 bis 4 zeigt sich ein uneinheitlicher Interaktionseffekt, welcher sich zwischen den betrachteten Varianzparameter unterscheidet. Besonders hervorzuheben ist, dass die Bedeutung der Invarianzformen für die Ko-/Varianzschätzungen der Wachstumsparameter über die Klassendistanz moderiert wird, was nicht zuletzt durch den durchgängig höheren Anteil bzgl. der Varianzaufklärung deutlich wird.

In der Gesamtbetrachtung lässt sich zusammenfassen, dass Fallzahl, Invarianzform, Klassen-separation und die Interaktion zwischen Invarianzform und Klassen-separation substantiell bedeutsame Einflüsse auf die Varianzparameterschätzungen ausüben. Dabei darf jedoch nicht vernachlässigt werden, dass der Zufallsprozess der Ziehung ebenfalls einen bedeutsamen Einfluss ausübt. Insgesamt zeigen die R^2 -Werte der Metamodelle mit einer Varianzaufklärung zwischen 27 und 35 % (vgl. Tabelle 20 auf Seite 132), dass große Anteile der Varianzschätzungen der – bzgl. der Klassenanzahl richtig identifizierten – Replikationen vom Zufallsprozess der Monte-Carlo-Simulation und Werten außerhalb des Analysemodells abhängen und nicht über die Simulationsgrößen erklärt werden können.⁶⁵

⁶⁵Dieses Ergebnis wurde durch vollsaturierte Metamodelle mit allen möglichen Interaktionen der Simu-

10.2.2 Analyse bzgl. des Bias der Interceptfaktor-Mittelwerte

Die Metaanalysemodelle bzgl. der Mittelwertverzerrungen der Interceptfaktoren der einzelnen latenten Klassen weisen deutliche Unterschiede zueinander auf. Der Grad der statistisch erklärten Varianz der Bias-Modelle schwankt zwischen unter 5 % bis hin zu ca. 28 %. Daher werden die Verzerrungsmodelle im Weiteren für die einzelnen latenten Klassen zunächst getrennt diskutiert.

- Die unabhängigen Variablen der Simulation haben nur einen äußerst geringen Einfluss auf die Mittelwertverzerrung des Interceptfaktors der ersten latenten Klasse (vgl. Tabelle 22: Modell 1 und Tabelle 23: c1). Die kumulierte Varianzaufklärung bleibt knapp unter 5 % zurück. Dementsprechend zeigen sich auch für keinen der spezifizierten Effekte bedeutsame Delta- R^2 -Werte. Den vergleichsweise stärksten Einfluss übt der Haupteffekt der Invarianzform aus, mit einem Delta- R^2 -Wert von 2.3 %. Zwar zeigt sich in der Tendenz anhand der Koeffizienten für höhere Fallzahl und höhere Klassendistanz unverzerrtere Parameterschätzung, aber vor dem Hintergrund der geringen Varianzaufklärung dürfen diese Tendenzen nicht als substantiell relevante Effekte überinterpretiert werden. Für die Invarianzformen zeigt sich – auf Grundlage der deskriptiven Analyse erwartbar – ein höherer Grad an Verzerrung für Konfigurationen, welche non-invariante Faktorladungen enthalten (3, 5, 6). Aber auch dieser Effekt bleibt beinahe unter der zuvor festgelegten Interpretationsgrenze zurück. Analytisch betrachtet erscheint der fehlende substantielle Einfluss der Simulationsgrößen auf Mittelwertverzerrungen des Interceptfaktors der ersten latenten Klasse zunächst kontraintuitiv. Die erste latente Klasse wird für die Invarianzformen sowohl in den längsschnittlichen als auch in den gruppenspezifischen Konfigurationen hinsichtlich der Nicht-Invarianz manipuliert. Diese Verzerrungen scheinen von den Interceptsschätzungen der dritten nicht-manipulierten Referenzklasse aufgefangen zu werden. Des Weiteren fallen die Anteilsverhältnisse der ersten Klasse auch nie unter einen kritischen Grenzwert (in Übereinstimmung mit den Vorgaben des Populationsmodells), was die fehlende Einflussbeziehung plausibilisiert.
- Das Mittelwertverzerrungsmodell für den Interceptfaktor der zweiten Klassen weist im Vergleich zur ersten latenten Klasse wesentlich stärkere Einflussbeziehungen auf

lationsgrößen (dabei bis zu Interaktionen 3. Ordnung) bestätigt, welche keine bedeutsam höheren Werte für R^2 aufweisen. Gleiche Tests wurden auch für alle weiteren diskutierten Modelle durchgeführt.

(vgl. Tabelle 22: Modell 2 und Tabelle 23: c2). Die kumulierte Varianzaufklärung beträgt immerhin mehr als 27 %. Davon geht der größte Anteil auf die Invarianzformen sowie die Interaktion zwischen Invarianzform und Klassenseparation zurück. Der zweitgrößte Einfluss geht vom Klassenanteilsverhältnis in Kombination mit der Klassendistanz aus, wobei die Haupteffekte der Klassenseparation und des Anteilsverhältnisses mit ca. zwei und einem Prozent zusätzlicher Varianzaufklärung auf Rang 4 und 5 stehen. Wie bisher bei allen betrachteten Metamodellen zeigt der Haupteffekt der Klassendistanz, dass für höhere Abstände zwischen den Klassen die Parameterschätzung im Schnitt weniger verzerrt wird. Der Haupteffekt der Invarianzformen zeigt für die gruppenspezifische und die kombinierte Messinvarianzstufen 5 und 6 eine deutlich höhere Verzerrung der Mittelwertschätzung des Interceptfaktors an. Bezieht man in diese Betrachtung den Interaktionseffekt zwischen den Invarianzformen und der Klassendistanz ein, lassen sich zwei weitere Tendenzen ausmachen. Für höhere Klassendistanzen werden längsschnittliche Invarianzformen, welche sowohl die Intercepts als auch die Faktorladungen umfassen, bedeutender und führen im Schnitt zu stärkeren Mittelwertverzerrungen des Interceptfaktors. Des Weiteren zeigt sich, dass für Klassendistanzen ab 1.5 MD der verzerrende Effekt von gruppenspezifischen Nicht-Invarianzen, welche die Intercepts und Faktorladungen betreffen, abnimmt. Die beiden Tendenzen überlagern sich in der gruppenspezifisch und längsschnittlich kombinierten Invarianzstufe 6, wodurch sich der Unterschied zwischen dem zusätzlichen Effektbeitrag für 1.5 und 4.0 MD potenziell erklärt. Der Haupteffekt der Klassenanteilsverhältnisse hat mit einem Delta- R^2 -Wert von 1.1 % einen verhältnismäßig geringen Einfluss und zeigt für Konfigurationen mit kleineren Klassen eine weniger verzerrte Schätzung des Mittelwerts des Interceptfaktors der zweiten Klasse. Bezieht man in dieser Betrachtung den Interaktionsterm zwischen den Klassenanteilsverhältnissen und der Klassendistanz mit ein, was einer zusätzlichen Varianzaufklärung von ca 5 % entspricht, dann zeigt sich, dass dieser Effekt stark konditional mit der Klassendistanzen verknüpft ist. Für eine Distanz von 1.5 MD lässt sich kaum eine Verringerung oder Erhöhung der Verzerrung ausmachen. Ab einer Klassendistanz von 4.0 MD zeigt sich aber, dass Modelle mit kleinen Klassengrößen von 5 % im Schnitt zu verzerrteren Interceptschätzungen in selbiger kleiner Klasse führen.

- Die Verzerrung des Interceptfaktor-Means der dritten latenten Klasse werden ebenfalls nicht unwesentlich durch die Simulationsgrößen bestimmt. Die kumulierte Varianzaufklärung fällt mit knapp 19 % (vgl. Tabelle 23: c3 auf Seite 140) zwar nicht so hoch aus wie für die zweite Klasse, liegt aber noch deutlich über dem Wert für das Metamodell bzgl. der Verzerrungen der Interceptsmeans der ersten latenten Klasse. Diese Varianzaufklärung wird im Wesentlichen durch die Invarianzstufen, die Klassendistanz und die Interaktion zwischen diesen beiden Simulationsgrößen erzeugt. Die Klassenanteilsverhältnisse spielen für die dritte Klasse wiederum keine wesentliche Rolle und auch die Fallzahl bleibt unter einer zusätzlichen Varianzaufklärung von einem Prozent. Für die Klassenseparation zeigt sich wiederum die erwartete Effektrichtung. Je höher die Distanz zwischen den latenten Klassen, desto unverzerrter wird der Mittelwert des Intercepts der dritten Klasse im Durchschnitt geschätzt (vgl. Tabelle 22: Modell 3 auf Seite 139). Trotz der kaum interpretationswürdigen Stärke der Assoziation, zeigen die Koeffizienten der Fallzahl in der Tendenz an, dass höhere Fallzahlen eher mit einer geringeren Verzerrung des Interceptmittelwerts einhergehen. Der Haupteffekt der Invarianzstufen zeigt für geringe Klassendistanzen bei gruppenspezifischen Nicht-Invarianzen geringere Verzerrungen. Unter Einbezug des Interaktionsterms zeigt sich allerdings, dass dieser Effekt bei höheren Klassendistanzen von 1.5 MD oder 4.0 MD nicht mehr vorhanden ist. In der höchsten Klassendistanz sind gruppenspezifische Nicht-Invarianzen sogar mit größeren Verzerrungen assoziiert. In diesem Interaktionseffekt spiegelt sich die (umgekehrte) Tendenz aus dem Modell für Klasse 2, welche für kleine Klassendistanzen zunächst eine verzerrtere Schätzung feststellt, was sich allerdings für höhere Klassendistanzen abschwächt.

In der Gesamtbetrachtung fällt auf, dass die Fallzahl für keinen der Mittelwerte der Interceptfaktoren einen bedeutenden Einfluss hat. Zwar zeigt die Tendenz der Effektkoeffizienten an, dass höhere Fallzahlen im Schnitt zu geringeren Verzerrungen führen, aber die damit verbundene zusätzliche Varianzaufklärung bleibt immer unter einem Prozent. Außerdem wird deutlich, dass die Simulationsgrößen nicht nur für diejenigen latenten Klassen einen relevanten Einfluss zeigen, in welchem tatsächlich auch im Populationsmodell modelltechnische Manipulationen durchgeführt wurden. So zeigt sich zwar der Effekt der Anteilverhältnisse nur in der zweiten Klasse, welche durch die Manipulation dieser Simulationsgröße vom Fallzahlverhältnis tatsächlich betroffen ist und kleine Klassengrößen aufweist, aber gleiches gilt nicht für den Einfluss der Messinvarianzformen, welche vor-

rangig für die Mittelwerte der manifesten Intercepts der Klassen 2 und 3 relevant werden. Dies ist insofern überraschend, als dass die gruppenspezifischen Faktorladungen und manifesten Interceptwerte im Rahmen der Simulation lediglich für die Klassen 1 und 2 manipuliert werden, aber der verzerrende Effekt von den Schätzungen durch Klasse 3 aufgefangen wird. Dabei zeigt sich zugleich, dass höhere Abweichungen von der Nicht-Invarianz nicht zwingend mit einer höheren Einflussstärke einhergehen, wenn man die Erklärungsanteile bzgl. Klasse 2 und Klasse 3 vergleicht.⁶⁶ Resümierend ist auch festzustellen, dass höhere Klassendistanzen durchweg mit einer geringeren Verzerrung der Mittelwerte der Interceptfaktoren assoziiert sind. Dabei sticht dieser Effekt allerdings für Klasse 1 und 2 hinsichtlich der relativen Bedeutung nicht besonders hervor. Alles in allem bleibt festzuhalten, dass die Interceptschätzungen sowohl von der Klassendistanz, den Invarianzformen und der Interaktion zwischen beiden Größen abhängen. Die Wirkung lässt sich dabei nicht klar an den durch die Simulation manipulierten Klassen festmachen, sondern strahlt zum Teil auf die Referenzklasse aus.

10.2.3 Analyse bzgl. des Bias der Slopefaktor-Mittelwerte

Während die Interceptparameter das Basisniveau der Wachstumsverlaufsform beschreiben, stehen die Slopeparameter für die zeitliche Veränderung des Verlaufs. Insofern beinhalten sie den dynamischen Aspekt der MIGMMs.

Die Metamodelle für die Verzerrungen der Slopemittelwerte zeigen im Vergleich zu den Modellen für die Varianzen und die Interceptwerte eine deutlich höhere Varianzaufklärung. Für die latenten Klassen 1 und 2 können ca. 47 % und 67 % statistisch erklärt werden, wohingegen für die Referenzklasse 3 immerhin noch 28 % der Varianz gebunden werden (vgl. Tabelle 25 auf Seite 143). Alle drei Analysemodelle haben gemeinsam, dass diese Varianzaufklärung weder wesentlich auf die Fallzahl, die Klassenanteilsverhältnisse noch auf den Haupteffekt der Invarianzstufen zurückzuführen ist. Die zugrundeliegenden Delta- R^2 -Werte bleiben alle deutlich unter einem Prozent zusätzlicher Varianzaufklärung zurück (vgl. Tabelle 24 auf Seite 142). So zeigen die unstandardisierten Effektschätzer für die Fallzahl zwar erwartbar an, dass höhere Fallzahlen zu geringeren Verzerrungen führen, aber der Effekt fällt insgesamt gering aus. Auch das Klassenanteilsverhältnis zeigt annahmekonform für die zweite latente Klasse eine höhere Verzerrung für geringe Klassenanteile

⁶⁶Die Abweichungen von der Nicht-Invarianz der manifesten Intercepts der Messfaktoren (der einzelnen Erhebungswellen) und der Faktorladungen fallen, nach der Definition der Populationsmodelle, für die latente Klasse 1 doppelt so stark aus wie für Klasse 2.

Tabelle 22: Regressions-Ergebnisse zum Bias der Interceptfaktoren (robuste SE; HC3)

	<i>Dependent variable:</i>		
	model_bias_i_m_c1	model_bias_i_m_c2	model_bias_i_m_c3
	(1)	(2)	(3)
Constant	0.432*** (0.032)	0.172*** (0.007)	0.421*** (0.018)
nobs2	-0.045 (0.031)	-0.013*** (0.002)	-0.035*** (0.004)
nobs3	-0.083** (0.028)	-0.021*** (0.002)	-0.056*** (0.004)
mixing2	-0.006 (0.021)	-0.042*** (0.008)	0.003 (0.003)
mixing3	-0.004 (0.025)	-0.083*** (0.008)	0.026*** (0.004)
invar2	0.067* (0.028)	0.002 (0.006)	-0.039 (0.024)
invar3	0.114*** (0.008)	0.016* (0.007)	-0.054* (0.024)
invar4	0.094*** (0.008)	0.010 (0.007)	-0.083*** (0.023)
invar5	0.526*** (0.028)	0.180*** (0.009)	-0.231*** (0.019)
invar6	0.571*** (0.035)	0.186*** (0.009)	-0.203*** (0.019)
sep2	-0.085* (0.039)	-0.036*** (0.008)	-0.181*** (0.023)
sep3	-0.339*** (0.026)	-0.132*** (0.007)	-0.366*** (0.017)
invar2:sep2		0.025** (0.008)	-0.001 (0.029)
invar3:sep2		0.026** (0.008)	0.011 (0.030)
invar4:sep2		-0.013 (0.009)	0.039 (0.028)
invar5:sep2		-0.189*** (0.010)	0.226*** (0.025)
invar6:sep2		-0.163*** (0.011)	0.245*** (0.026)
invar2:sep3		0.011 (0.007)	0.052* (0.024)
invar3:sep3		0.033*** (0.007)	0.060* (0.024)
invar4:sep3		0.111*** (0.008)	0.125*** (0.023)
invar5:sep3		-0.123*** (0.009)	0.315*** (0.019)
invar6:sep3		-0.073*** (0.010)	0.267*** (0.020)
mixing2:sep2		0.039*** (0.008)	
mixing3:sep2		0.052*** (0.008)	
mixing2:sep3		0.066*** (0.008)	
mixing3:sep3		0.175*** (0.008)	
Observations	23,084	23,084	23,084
R ²	0.046	0.276	0.186
Adjusted R ²	0.045	0.275	0.186
Residual Std. Error	1.386 (df = 23072)	0.117 (df = 23058)	0.214 (df = 23062)
F Statistic	100.175*** (df = 11; 23072)	351.858*** (df = 25; 23058)	251.675*** (df = 21; 23062)

Info:

*p<0.05; **p<0.01; ***p<0.001

Tabelle 23: Delta-R-Quadrat der Praediktoren zum Bias der Intercept Parameter

	df	dR Mean I c1	dR Mean I c2	dR Mean I c3
nobs	2	0.001	0.004	0.009
mixing	2	0.00000	0.011	0.002
invar	5	0.023	0.079	0.027
sep	2	0.010	0.018	0.049
invar:sep	10		0.074	0.040
mixing:sep	4		0.048	

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 22.

an, aber auch dieser Effekt kann mit einem Delta- R^2 von unter einem Prozent kaum als bedeutsam eingestuft werden.

Das Gros der Varianzbindung geht auf den Interaktionseffekt zwischen der Klassendistanz und den Invarianzstufen sowie in einem geringeren Umfang auf die Klassendistanz selbst zurück. Die Klassendistanz ist – nicht überraschend – bei großer Klassenseparation mit geringeren Verzerrungen der Slopemittelwerte assoziiert. Dieser Effekt ist besonders für die dritte Referenzklasse bedeutsam, wohingegen der Effekt für die Klassen 1 und 2 geringer, aber dennoch substantiell ausfällt. In der Gesamtbetrachtung für den Haupt- und Interaktionseffekt der Invarianzstufen zeigt sich, dass die Slopes gegenüber den längsschnittlichen Verzerrungen relativ robust sind. Nur für die Referenzklasse 3 zeigen sich bei großer Klassendistanz verzerrende Effekte für die längsschnittlichen Simulationskonfigurationen, was die zuvor dargelegten deskriptiven Auswertungen zum relativen Bias bestätigt. Für die gruppenspezifischen und kombinierten Invarianzkonfigurationen gestaltet sich das Bild wiederum etwas differenzierter, weshalb eine getrennte Betrachtung sinnvoll ist.

- Für den Mittelwert der Slopes der ersten latenten Klasse zeigt sich bei geringer Klassendistanz sogar eine geringere Verzerrung für die Invarianzstufen 5 und 6 (vgl. Tabelle 25: Modell 1 auf Seite 143). Diese Tendenz dreht sich allerdings bereits bei einer mittleren Klassendistanz von 1.5 MD. Hier zeigt sich im Schnitt eine größere Verzerrung für die Invarianzstufen 5 und 6 und im geringeren Umfang für die Invarianzstufe 4. Der Grad der Verzerrung wird in der höchsten Klassendistanz von 4.0 MD noch deutlich avancierter und steigt im Vergleich zur mittleren Distanz um den Faktor 5 bis 10 an. Die kombinierte Invarianzstufe 6 fällt hierbei etwas weniger verzerrt aus als Stufe 4 und 5, weshalb hier von einer gewissen Kompensationsleistung der Kombination von längsschnittlicher und gruppenspezifischer Nicht-Invarianz gesprochen werden kann.

- Die Mittelwerte der zweiten latenten Klasse sind für geringe Klassendistanzen, im Gegensatz zur ersten latenten Klasse, zunächst verzerrter geschätzt (vgl. Tabelle 25: Modell 2 auf Seite 143). Für die mittlere Klassendistanz dreht sich diese Beziehung (zumindest für nicht-invariante Faktorladungen) wieder, womit die Schätzungen sogar leicht weniger verzerrt sind. Aber auch hier zeigen sich in der größten Klassendistanz deutliche Verzerrungen, welche die der geringeren Klassendistanzen deutlich übersteigen. Auch hier fällt der Bias für die kombinierte Invarianzstufe etwas geringer aus, was wiederum eine kleine Kompensationsleistung vermuten lässt, ohne das dadurch allerdings akzeptable Verzerrungsraten erreicht werden.
- Mit einer vergleichsweise geringen zusätzlichen Varianzaufklärung der Invarianzstufen und des Interaktionseffekts fallen die Effekte für die dritte latente Klasse deutlich weniger bedeutsam aus, wie für die anderen Klassen. Dennoch zeigen sich einige verzerrende Effekte für die Invarianzstufen, welche durch nicht-invariante Spezifikationen in den anderen beiden latenten Klassen bedingt werden. In der geringen Klassendistanz zeigt sich beispielsweise, dass für gruppenspezifische Nicht-Invarianz der manifesten Intercepts der Slopemittelwert verzerrter geschätzt wird als für größere Klassendistanzen (vgl. Tabelle 25: Modell 3 auf Seite 143). Für eine mittlere Klassendistanz zeigen sich bzgl. der Stärke gleichgearteter Effekte für die Invarianzstufen mit gruppenspezifischen Nicht-Invarianzen in den Faktorladungen (5, 6), wobei die 4. Stufe weder weniger noch mehr verzerrt geschätzt wird. Für die höchste Klassendistanz nehmen die Verzerrungen für die Invarianzstufen 5 und 6 um ca. $\frac{2}{3}$ ab und für die 4. Invarianzstufe lässt sich sogar eine weniger verzerrte Schätzung des Slopemittelwerts identifizieren.

Während die Verzerrungen bei längsschnittlichen Nicht-Invarianzstufen eher gering ausfallen, werden die Slope-Mittelwerte vor allem durch die gruppenspezifischen und die kombinierten Konfigurationen beeinflusst. Zusammenfassend betrachtet, werden vorrangig die Slopeschätzungen der manipulierten Klassen durch die gruppenspezifischen Invarianzstufen beeinflusst. Für sie zeigen sich insbesondere bei klarer Klassentrennung starke Verzerrungen der Mittelwertschätzungen. Auch die Richtung der manipulierten Nicht-Invarianzen scheint insbesondere bei kleinen Klassendistanzen eine Rolle zu spielen. Nichtsdestotrotz werden auch die Slopeschätzungen der Referenzklasse 3 beeinflusst, wobei dem Haupteffekt der Klassendistanz hier eine größere Relevanz zukommt. Eine

alleinige Verzerrung der durch das Populationsmodell nicht-invariant gesetzten latenten Klassen trifft, wie auch schon bei den Intercept-Metamodellen, nicht gänzlich zu. Auch für die Slopemodelle gibt es Ausstrahlungseffekte auf die Nachbarklassen, wenn auch deutlich geringer, als dies bei den Intercept-Metamodellen der Fall war. Neben den Invarianzformen stellt die Klassendistanz insgesamt die einzige weitere bedeutsame Einflussgröße für die Verzerrungen der Slopemittelwerte dar. Resümierend lässt sich auch sagen, dass sich die Verzerrungen der Invarianzstufen besonders dann herausstellen, wenn die Distanz zwischen den latenten Klassen besonders groß ist, was eine Fehlzuordnung der Fälle innerhalb der latenten Klassen deutlich unwahrscheinlicher macht.

Tabelle 24: Delta-R-Quadrat der Praediktoren zum Bias der Slope Parameter

	df	dR Mean S c1	dR Mean S c2	dR Mean S c3
nobs	2	0.0002	0.0001	0.006
mixing	2	0.0002	0.005	0.001
invar	5	0.001	0.001	0.004
sep	2	0.011	0.010	0.030
invar:sep	10	0.179	0.231	0.024

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 25.

Tabelle 25: Regressions-Ergebnisse zum Bias der Slopefaktoren (robuste SE; HC3)

	<i>Dependent variable:</i>		
	model_bias_s_m_c1	model_bias_s_m_c2	model_bias_s_m_c3
	(1)	(2)	(3)
Constant	0.216*** (0.009)	0.122*** (0.006)	0.210*** (0.008)
nobs2	-0.008* (0.004)	-0.006* (0.003)	-0.019*** (0.002)
nobs3	-0.010** (0.004)	-0.006* (0.003)	-0.031*** (0.002)
mixing2	-0.009*** (0.003)	0.024*** (0.002)	-0.001 (0.002)
mixing3	0.002 (0.004)	0.049*** (0.003)	0.011*** (0.002)
invar2	0.009 (0.013)	0.016* (0.008)	0.010 (0.011)
invar3	0.018 (0.013)	0.025** (0.008)	-0.016 (0.010)
invar4	0.027* (0.013)	0.022* (0.009)	0.069*** (0.012)
invar5	-0.042*** (0.010)	0.042*** (0.008)	-0.012 (0.009)
invar6	-0.031** (0.010)	0.059*** (0.008)	0.0002 (0.010)
sep2	-0.010 (0.013)	0.084*** (0.010)	-0.058*** (0.012)
sep3	-0.190*** (0.009)	-0.098*** (0.006)	-0.181*** (0.008)
invar2:sep2	-0.012 (0.017)	-0.011 (0.012)	-0.025 (0.015)
invar3:sep2	0.037 (0.019)	-0.043*** (0.012)	-0.012 (0.015)
invar4:sep2	0.064*** (0.019)	0.050*** (0.012)	-0.017 (0.016)
invar5:sep2	0.099*** (0.015)	-0.075*** (0.012)	0.092*** (0.013)
invar6:sep2	0.093*** (0.015)	-0.098*** (0.012)	0.087*** (0.014)
invar2:sep3	-0.002 (0.013)	-0.006 (0.008)	0.023* (0.011)
invar3:sep3	0.035** (0.013)	0.001 (0.009)	0.046*** (0.010)
invar4:sep3	0.630*** (0.014)	0.677*** (0.011)	-0.070*** (0.012)
invar5:sep3	0.627*** (0.015)	0.438*** (0.008)	0.034*** (0.009)
invar6:sep3	0.500*** (0.016)	0.345*** (0.010)	0.035*** (0.010)
Observations	23,084	23,084	23,084
R ²	0.467	0.669	0.278
Adjusted R ²	0.467	0.669	0.277
Residual Std. Error (df = 23062)	0.227	0.152	0.134
F Statistic (df = 21; 23062)	962.968***	2,224.196***	422.036***

Info:

*p<0.05; **p<0.01; ***p<0.001

10.3 Diskussion zu den modellbasierten Analysen bzgl. des Bias

Die vorgestellten Ergebnisse zu den Bias-Werten der Wachstumsfaktoren müssen vor dem Hintergrund der vorangestellten korrekten Identifikation der Klassenanzahl bewertet werden. Alle Aspekte des Bias, welche durch den Prozess der Klassenidentifikation bedingt sind⁶⁷, konnten folglich auch in den Analysemodellen nicht berücksichtigt werden. Aufgrund der praktischen Verfahrenslogik bei der Modellkonstruktion von (MI)-GMMs, welche auch in Einführungs- und Lehrtexten als konsekutiv beschrieben wird, erscheint die alleinige Betrachtung der korrekt identifizierten Modelle als folgerichtig und für die praktische Anwendung in der Empirie am sinnvollsten. Aussagen zu den Verzerrungen der Wachstumsverläufe für alle 3-Klassenmodelle, unabhängig von der Klassenidentifikation, wären zwar prinzipiell möglich gewesen, wären aber bzgl. der Empfehlungen für die Forschungspraxis im besten Fall irrelevant bis hin zu potentiell irreführend.

Betrachtet man die Erklärungskraft der Bias-Metamodelle der verschiedenen Parameter der Wachstumsfaktoren kann man erkennen, dass die Varianz des Bias bzgl. der Ko-/Varianzen und der Mittelwerte der Intercept- und Slope-Faktoren unterschiedlich gut durch die Simulationsgrößen statistisch erklärt werden kann. Während sich die R^2 -Werte für die Ko-/Varianz im Bereich von 27 bis 35 % bewegen, zeigen sich für die Mittelwerte der Interceptfaktoren lediglich Werte von 5 bis 28 %. Die Unterschiedlichkeit der Mittelwertverzerrungen der Slopefaktoren kann hingegen mit R^2 -Werten im Bereich von 28 bis 67 % am besten gebunden werden. Diese Unterschiedlichkeit in der Erklärungskraft geht tlw. auch mit unterschiedlichen Effekten der Simulationsgrößen einher.

Während der Fallzahl der Simulation für die Varianzschätzungen z. B. eine verhältnismäßig hohe Relevanz zukommt, spielt sie für die Verzerrungen der Mittelwerte der Intercept- und Slopefaktoren kaum eine nennenswerte Rolle. Auch für die Klassenanteilsverhältnisse zeigt sich ein unterschiedlich gearteter Effekt über die Metamodelle hinweg. So haben die Klassenanteilsverhältnisse lediglich für den Bias des Interceptfaktors der zweiten latenten Klasse einen nennenswerten Einfluss, wohingegen sie in den restlichen Analysemodellen kaum zusätzliche Varianz binden können. Andere Simulationsgrößen haben hingegen durchgehend über alle Metamodelle hinweg einen Einfluss auf die Bias-Werte. Eine höhere Klassendistanz hat in jedem Analysemodell einen positiven Einfluss auf die Bias-Werte und vermag in nahezu jedem⁶⁸ dieser Modelle einen eigenen substantiellen Beitrag zur

⁶⁷Dabei ist sowohl an Haupt- wie auch an mögliche Interaktionseffekte zu denken.

⁶⁸Das Metamodell zum Bias des Intercept-Mittelwerts der ersten latenten Klasse könnte als Ausnahme gelten, wobei allerdings zu beachten ist, dass die Gesamtvarianzaufklärung des Modells äußerst gering

Varianzaufklärung beizusteuern. Ähnlich verhält es sich mit den Invarianzstufen und insbesondere mit der Interaktion zwischen den Invarianzstufen und der Klassendistanz. Für die längsschnittlichen Invarianzformen kann anhand der deskriptiven sowie der modellbasierten Auswertungen gezeigt werden, dass sie im Vergleich zu den gruppenspezifischen Konfigurationen deutlich weniger zu den Bias-Werten beitragen. Wenn allerdings verzerrende Effekte der längsschnittlichen Invarianz auftreten, dann in der dritten Kategorie (invar3), welche nicht nur zeitlich nicht-invariante manifeste Intercepts beinhaltet, sondern zugleich nicht-invariant spezifizierte Faktorladungen über die Erhebungszeitpunkte hinweg.

Die Effekte fehlender gruppenspezifischer Nicht-Invarianz fallen im Schnitt stärker aus als die längsschnittlichen Abstufungen. Während eine alleinige nicht-invarianz der manifesten Intercepts für die Ko-/Varianzschätzungen noch keinen Effekt aufweist, zeigen sich deutliche Verzerrungen bei der Intercept- und Slope-Faktoren. Bei zusätzlicher nicht-invarianz der Faktorladungen zeigen sich für alle geschätzten Koeffizienten Verzerrungen. Für die Ko-/Varianzschätzungen führt dies durchgehend zu einer Erhöhung der Verzerrung, während sich für die Intercept- und Slope-Faktoren differenzierte Wirkungen ausmachen lassen. Für die Slopefaktoren führt eine bzgl. des Vorzeichens gleichgerichtete Nicht-Invarianz der Faktorladungen zu einer Verringerung der Verzerrung der Mittelwertschätzungen. Für die Interceptfaktoren lässt sich, abhängig von der betrachteten latenten Klassen, sowohl eine Verstärkung wie auch eine Verringerung identifizieren.

Für die kombinierte Invarianzkonfiguration, welche längsschnittliche und gruppenspezifische Nicht-Invarianz kombiniert, werden die Ko-/Varianzschätzungen im Vergleich zu den gruppenspezifischen Ausprägungen sogar stärker verzerrt geschätzt. Folglich verstärkt sich auch der Bias bzgl. der Heterogenitätseinschätzungen, wenn beide Formen der Nicht-Invarianz kombiniert auftreten. Ähnlich verhält es sich für die Mittelwertschätzungen der Interceptfaktoren. Allerdings bleibt die verzerrende Wirkung hinter der einfachen Addition der Verzerrungen der gruppenspezifischen und längsschnittlichen Konfigurationen zurück. Im Gegensatz hierzu führt die Kombination der beiden Invarianzformen für die Schätzung der Mittelwerte der Slopefaktoren zu geringerem Bias. Diese Verringerung oder Kompensation des Bias fällt allerdings keineswegs so stark aus, als dass vom Fehlen von Bias gesprochen werden kann.⁶⁹

ausfällt und die Klassenseparation, in Anbetracht dessen, verhältnismäßig gut abschneidet.

⁶⁹Die Reduzierung im Vergleich zur gruppenspezifischen Ausprägung bewegt sich nach den Metamodellschätzungen (für die größte Klassendistanz) etwa im Bereich von 20 %.

Zugleich wird dieser Kompensationseffekt durch eine höhere Verzerrung in der nicht manipulierten Referenzgruppe erkaufte. Insbesondere für die Mittelwertverzerrungen der Interceptfaktoren, aber auch in einem geringeren Umfang für die Slopefaktormittelwerte, konnte gezeigt werden, dass Verzerrungen nicht nur in den von der Nicht-Invarianz betroffenen latenten Gruppe auftreten, sondern auch Ausstrahlungseffekte auf Nachbarklassen zu erwarten sind, was in Folge dessen auch deren inhaltliche Interpretation kompromittiert.

Über die Modelle hinweg lässt sich der Trend erkennen, dass die Verzerrungen durch die Invarianzformen erst bei einer großen Klassendistanz voll zum Tragen kommen, was nicht zuletzt auch die hohen relativen Bias-Werte der deskriptiven Tabellenanalysen plausibilisiert (bei welchen die Klassendistanz auf 4.0 MD fixiert wurde). Anhand des einheitlichen Musters dieser Ergebnisse liegt die Vermutung nahe, dass bei großer Klassendistanz die Abweichungen durch die Invarianzenverletzungen nicht mehr ausreichend durch eine «falsche» Zuordnung der Fälle in benachbarte latente Klassen abgefangen werden kann und sich somit stärker in der jeweiligen Klasse bzgl. des Bias auswirkt.⁷⁰ Zhang 2016 kommt, allerdings nur auf Grundlage von längsschnittlichen Nicht-Invarianzformen, zu einem gleichen Ergebnis bzgl. der Interaktion zwischen Invarianzformen, Klassendistanz und Bias-Werten (vgl. Zhang, 2016: 104).

⁷⁰Dieser Aspekt wird auch in Abschnitt 10.4 ausführlicher thematisiert.

Tabelle 26: Zusammenfassung der Bias-Performanz bzgl. der Wachstumsfaktoren

	Kovarianz/Varianzen	Intercept-Mittelwerte	Slope-Mittelwerte
	unbedeutend (für FL vorhanden)	verzerrt (für FL)	unbedeutend (Ausnahme Referenzklasse)
Fehlende Messinvarianz	verzerrt (insb. bei hoher Klassendistanz)	verzerrt	verzerrt (bei hoher Klassendistanzen)
	stark verzerrt (insb. bei hoher Klassendistanz)	verzerrt	geringfügig verzerrt (bei hoher Klassendistanz)
	weniger Verzerrung	weniger Verzerrung	weniger Verzerrung
Hohe Klassendistanz	weniger Verzerrung	unbedeutend	unbedeutend
Hohe Fallzahl	unbedeutend	nur betroffene Klasse verzerrt	unbedeutend
Kleine Klasse	27 bis 35 % nicht sinnvoll identifizierbar	5 bis 28 % stark vorhanden	28 bis 67 % vorhanden

10.4 Exkurs: Bias der geschätzten Klassenanteile

Um den Effekt falscher Klassenzuordnung genauer zu betrachten, wurde Ad hoc eine weitere Maßzahl aus den Simulationsergebnissen erstellt, welche den Grad der Verzerrung bzgl. der Klassenzuordnung (zumindest im Ansatz) abbilden kann. Die prozentuale Klassenverzerrung ist ein Maß dafür, wie weit die Klassenanteile vom definierten Wert des Populationsmodells (über die drei Klassen) gemittelt abweichen.⁷¹ Dabei handelt es sich allerdings nur um ein proximales Maß für die falsche Klasseneinteilung, da auch eine Deckungsgleichheit der Prozentzahlen in den Klassen bei einer völlig falschen Klassenzuteilung prinzipiell möglich ist. Umgekehrt zeigt allerdings eine Abweichung von den Prozentzahlen des Populationsmodells in jedem Fall eine falsche Klassenzuordnung an.⁷² Erwartungsgemäß sind die Bias-Werte der dritten (nicht-manipulierten) Klasse bzgl. des Intercept- und Slope-Mittelwertes stark mit dem Grad der prozentualen Klassenverzerrung assoziiert (vgl. Tabelle 27 auf Seite 150). Dies kann als Hinweis darauf gedeutet werden, dass die Verzerrungen in der dritten Klasse maßgeblich durch die falsche Klassenzuordnung verursacht werden. Diese Assoziation untermauert damit die Interpretation von «Ausstrahlungseffekten» in den Kapiteln 10.2.2 und 10.2.3. Andersherum kann gezeigt werden, dass die prozentuale Klassenverzerrung eine Funktion der Simulationsgrößen darstellt. Ca. 74 % der Varianz der prozentualen Klassenverzerrung kann durch die Simulationsgrößen statistisch erklärt werden (vgl. Tabelle 28 auf Seite 151). Hohe Klassendistanz, Messinvarianz (oder lediglich längsschnittliche Non-Invarianz) und prozentual gleichverteilte Klassen begünstigen eine geringere prozentuale Klassenverzerrung, wobei die Klassendistanz den stärksten Einfluss hat. Die Fallzahl hat hingegen keinen eigenen Effekt auf die Verzerrung der Klassenanteile. Des Weiteren zeigen die Interaktionseffekte, dass verzerrende Effekte von kleinen Klassengrößen bei großen Klassendistanzen deutlich weniger stark ausfallen. Für die Messinvarianzstufen zeigt sich hingegen, dass die prozentualen Klassenverzerrungen für die gruppenspezifischen und die kombinierten Invarianzformen bei hoher Klassendistanz sogar stärker ausfallen. Folglich beeinflussen die Simulationsgrößen die (prozentuale) Klassenzuteilung und die verzerrte Klassenzuteilung

⁷¹Die prozentualen Klassenanteile variieren natürlich in Abhängigkeit von der Simulationsgröße «Mixing». Bei der Berechnung der Abweichung wurden die jeweiligen prozentualen Klassenteile je nach Simulationskonfiguration berücksichtigt. Für den Fall «Mixing=1» wurde beispielsweise die Abweichung vom Verteilungsmuster 33:33:33 berechnet. Für einen (fiktiven) Fall der Klassenverteilung von 40:30:30 würde für die prozentuale Klassenverzerrung ein Wert von $\frac{(7 + 3 + 3)}{3} = 4.3\bar{3}$ berechnet werden.

⁷²Ein weniger proximales Maß war leider aufgrund technischer Beschränkungen der Softwarepakete nicht (vor allem im Nachhinein) möglich.

ist mit Parameterverzerrungen der nicht-manipulierten latenten Klassen assoziiert. Dies plausibilisiert tlw. den Wirkmechanismus der Parameterbiasgenese für die Intercept- und Slopemittelwerte.

Ein expliziter Vergleich zwischen den Meta-Modellen zur Bias-Schätzung der Intercept- und Slopemittelwerte mit und ohne Kontrolle der Verzerrung der geschätzten Klassenanteile zeigt, dass die Einflüsse der Haupteffekte der Messinvarianzstufen, der Klassendistanz sowie deren Interaktion variiert.⁷³ Die Haupteffekte der Klassendistanz werden bei statistischer Kontrolle des Anteils falscher Klassenzuteilung allgemein geringer (auch hinsichtlich der Delta- R^2 -Werte). Die Haupteffekte der gruppenspezifischen Messvarianz werden hingegen stärker, da deren verzerrende Wirkung nicht mehr durch die Varianzanteile der falschen Klasseneinteilung reduziert wird. Der Einfluss der Interaktion zwischen der Klassendistanz und den Invarianzverletzungen nimmt hingegen ab, was inhaltlich so zu interpretieren ist, dass der verzerrende Effekt der Invarianzstufen nicht mehr so stark durch die Klassendistanz moderiert wird, da die Differenzen hinsichtlich der falschen Klasseneinteilung statistisch kontrolliert sind. Berücksichtigt man neben den Modellen mit und ohne Kontrolle der durchschnittlichen Klassenverzerrung zusätzlich das Modell aus Tabelle 28, wird deutlich, dass die falsche Klasseneinteilung als partieller Moderator für den Effekt der Simulationsvariablen auf die Verzerrungen der Intercept- und Slopemittelwerte fungiert (bzw. als Suppressor hinsichtlich des Haupteffekts der Invarianzstufen).

Einschränkend muss erwähnt werden, dass die durchschnittliche Klassenverzerrung nur als proximales Maß für eine falsche Einteilung von individuellen Fällen zu verstehen ist, welches aufgrund der Ex-post-facto-Natur dieser zusätzlichen Analyse nicht anders operationalisiert werden konnte. Die Maßzahl bewegt sich weder auf der Individualebene des einzelnen Falls, noch ist sie nach den drei latenten Klassen differenziert. Des Weiteren zeigt diese Zusatzanalyse, dass zwar Teile der Effekte der Simulationsgrößen auf den Bias durch falsche Klasseneinteilung erklärt werden können, aber da sich die Effekte unter Kontrolle der Klassenverzerrung nur abschwächen und nicht verschwinden, existieren noch weitere Wirkmechanismen.

⁷³Die Modelle für die folgenden Ausführungen entlang einer expliziten Mediatoranalyse finden sich im Anhang A.3. Dort sind die bereits diskutierten Meta-Modelle zum Bias unter zusätzlicher Berücksichtigung der durchschnittlichen Klassenverzerrung aufgeführt.

Tabelle 27: Korrelationsmatrix Bias-Werte

	Bias_v_i	Bias_v_s	Bias_cv_is	Bias_m_i_c1	Bias_m_i_c2	Bias_m_i_c3	Bias_m_s_c1	Bias_m_s_c2	Bias_m_s_c3
Bias_v_i									
Bias_v_s	0.25***								
Bias_cv_is	0.29***	0.28***							
Bias_m_i_c1	0.11***	0.06***	0.11***						
Bias_m_i_c2	0.26***	0.03***	-0.04***	0.17***					
Bias_m_i_c3	0.23***	0.14***	0.25***	0.04***	-0.01 n.s.				
Bias_m_s_c1	0.06***	0.02**	0.08***	0.08***	0.17***	-0.01 n.s.			
Bias_m_s_c2	-0.02***	-0.02***	-0.05***	0.01 n.s.	0.24***	-0.03***	0.65***		
Bias_m_s_c3	0.11***	0.30***	0.32***	0.05***	-0.02**	0.42***	-0.06***	-0.05***	
Bias_Klassenanteile	0.27***	0.35***	0.35***	0.20***	0.19***	0.43***	0.23***	0.10***	0.44***
Info: ***p<0.001; **p<0.01; *p<0.05;									
v=Varianz; cv=Kovarianz; i=intercept; s=slope; m=Mean der Klasse c;									
Fallzahl= 23084.									

Tabelle 28: Reg.-Ergebnisse zu durchschnittlichen Klassenverzerrungen (robuste SE; HC3)

	<i>Dependent variable:</i>	
	prob_mean_delta	
	Durchschnittliche Klassenverzerrung	
Constant	0.206*** (0.004)	
nobs2	−0.003* (0.001)	
nobs3	−0.003* (0.001)	
mixing2	0.034*** (0.003)	
mixing3	0.107*** (0.003)	
invar2	−0.002 (0.005)	
invar3	−0.015** (0.005)	
invar4	−0.004 (0.005)	
invar5	0.056*** (0.005)	
invar6	0.092*** (0.005)	
sep2	−0.068*** (0.005)	
sep3	−0.209*** (0.004)	
mixing2:sep2	0.010** (0.004)	
mixing3:sep2	0.010* (0.004)	
mixing2:sep3	−0.015*** (0.003)	
mixing3:sep3	−0.057*** (0.004)	
invar2:sep2	−0.005 (0.007)	
invar3:sep2	0.010 (0.006)	
invar4:sep2	−0.018** (0.007)	
invar5:sep2	0.030*** (0.006)	
invar6:sep2	0.061*** (0.007)	
invar2:sep3	0.003 (0.005)	
invar3:sep3	0.021*** (0.005)	
invar4:sep3	0.099*** (0.006)	
invar5:sep3	0.109*** (0.005)	
invar6:sep3	0.079*** (0.006)	
Observations	23,084	
R ²	0.738	
Adjusted R ²	0.737	
Residual Std. Error	0.071 (df = 23058)	
F Statistic	2,591.526*** (df = 25; 23058)	
Info:	*p<0.05; **p<0.01; ***p<0.001	

Tabelle 29: Delta-R-Quadrat der Praediktoren der geschätzten Klassenteilsverzerrungen

	df	dR-sqr
nobs	2	0.0001
mixing	2	0.020
invar	5	0.018
sep	2	0.042
mixing:sep	4	0.009
invar:sep	10	0.030

Info: Berechnungsgrundlage ist das Modell aus Tabelle 28.

11 Ergebnisse bzgl. der Standardfehler

Neben der Klassenidentifikation und der Verzerrung der Parameterschätzungen der Wachstumsverlaufsform existiert noch eine dritte Performanzgröße, welche für die Interpretation von Verlaufsformen von (MI)GMMs relevant ist. Die verzerrungsfreie Schätzung der Standardfehler ist für den Inferenzschluss auf die Grundgesamtheit und damit für die Verallgemeinerbarkeit der Ergebnisse von essenzieller Bedeutung. Da die Wachstumsverlaufsform im Rahmen von (MI)GMMs in aller Regel die inhaltlich relevante Größe darstellt, werden wiederum nur die Standardfehlerschätzungen der Wachstumsverlaufsform betrachtet. Zu große oder zu kleine Standardfehler führen zu breiteren oder schmalen Konfidenzintervallen, wodurch der Fehler erster und zweiter Art bei der Interpretation von empirischen Ergebnissen potenziell ungünstig beeinflusst wird.

Auch wenn in der empirischen Praxis kleine Standardfehler mit einer höheren Sicherheit bzgl. der Parameterschätzungen assoziiert sind und damit vordergründig als wünschenswert erscheinen, können Standardfehler auch zu klein ausfallen und damit ein falsches Niveau an Sicherheit bzgl. der Variabilität einer Parameterschätzung vermitteln (vgl. Carsey & Harden, 2014: 89).

Für die Bewertung von Standardfehlern werden in der Simulationsliteratur mehrere Maße vorgeschlagen. Hierzu zählen u. a. die Coverage Probabilities (CP), die eng mit den CPs verbundenen Rejection Rates und die Standard Deviation Method. Alle Maße außer der Standard Deviation Method beinhalten in ihrer Berechnung und Interpretationslogik neben der Verzerrung des Standardfehlers zusätzlich die Verzerrung des dazugehörigen Parameterschätzers. Dadurch können «Bias» und «SE-Bias» nicht getrennt voneinander betrachtet werden. Da bereits im vorangegangenen Abschnitt die Verzerrungen der Wachstumsparameterschätzer umfassend diskutiert wurden, sollen im Weiteren ebenfalls nur die Standardfehler betrachtet werden, weshalb für die weiteren Analysen die Standard Deviation Method als zugrundeliegendes Maß herangezogen wird.

Nach der Standard Deviation Method werden (simulations-)empirisch zu ermittelnde Standardabweichungen der Parameterschätzwerte als Quasi-Populationswerte für den Standardfehler herangezogen. Da die wiederholten Replikationen der Simulationsstudie der Konzeption der wiederholten Stichprobenziehung der Standardfehler entsprechen, sollten die berechneten Parameterschätzungen der Replikationen näherungsweise der Variabilität bei wiederholten Stichprobenziehungen entsprechen. Folglich entspricht die Standardab-

weichung der durch die Replikationen erzeugten Parameterschätzwerte einer adäquaten empirischen Schätzung des Populationsstandardfehlers, welcher für jeden Schätzparameter in jeder Simulationskonfiguration getrennt zu ermitteln ist.⁷⁴ Da es sich bei den Replikationen der Simulation nur um endlich viele «Stichproben» – und nicht um unendlich viele Stichproben wie nach der Logik des Standardfehlers – handelt, muss ein gewisses Maß an Schätzungenauigkeit des SE-Populationswertes hingenommen werden. Auf Grundlage des anhand der Replikationen ermittelten SE-Populationswertes können im nächsten Schritt individuelle Abweichungen zwischen den SE-Schätzwerten der Replikationen und dem SE-Populationswert ermittelt werden, um den SE-Bias zu berechnen. Auf Ebene der 162 Simulationskonfigurationen kann der relative SE-Bias (analog zum relativen Bias der Parameterschätzungen) als Maß für die durchschnittliche Abweichung und als Vergleichsmaß eingesetzt werden. Die Berechnung des relativen (skalenunabhängigen) SE-Bias gleicht der Formel des relativen Bias der Wachstumsparameterschätzungen:

$$100 * \frac{(\overline{SE - Schaetzwert} - SE\text{-Populationswert})}{SE\text{-Populationswert}}, \quad (19)$$

wobei der SE-Populationswert für jede Simulationskonfiguration einzeln geschätzt wird. Wie bereits bei der Analyse der Verzerrungen der Parameterschätzungen werden in den weiteren Analysen lediglich diejenigen Replikationen betrachten, welche zumindest von einem der IC-Indizes korrekt als 3-Klassenlösung identifiziert wurden.⁷⁵ Die inhaltliche Begründung für dieses Vorgehen kann aus dem entsprechenden vorangegangenen Kapitel 10 entnommen werden.

11.1 Haupteffekte bzgl. SE-Verzerrungen

Die folgenden deskriptiven Tabellenauswertungen zeigen die jeweils isolierten Haupteffekte der Simulationsgrößen auf den relativen SE-Bias.⁷⁶ Die jeweils nicht betrachteten Simulationsgrößen werden für die tabellarischen Auswertungen auf für die Schätzung möglichst optimale Simulationswerte fixiert.⁷⁷ In der Literatur wird für den relativen SE-Bias ein

⁷⁴Diese Form des geschätzten SE-Populationswertes wird auch als «empirischer Standardfehler» bezeichnet (vgl. Bandalos & Gagne, 2012: 101).

⁷⁵Für die Schätzung des SE-Populationswertes wurden hingegen jeweils alle Replikationen – unabhängig von der Klassenidentifikation – herangezogen. Extreme Parameterschätzungs-Ausreiser (> 15 Standardabweichungen) wurden für die Berechnung nicht berücksichtigt.

⁷⁶Eine vollständige Referenz zum SE-Bias findet sich im Anhang in den Tabellen 55 und 56.

⁷⁷Wie bereits in den vorangegangenen Kapiteln handelt es sich bei den möglichst «optimalen Werten» um die Ausprägungen: Invarianz der Messung (Invar=1), 1300 Fälle (Nobs=3), Gleichverteilung über die

Grenzwert von ± 0.10 bzw. 10 % vorgeschlagen (vgl. Hoogland & Boomsma, 1998: 346 ff.; Bandalos & Leite, 2013: 643 f.; Bandalos & Gagne, 2012: 101). Diese im Vergleich zum relativen Bias weichere Grenzwertempfehlung ist darauf zurückzuführen, dass für die Berechnung nicht auf einen Populationswert aus dem vorab definierten DGP zurückgegriffen werden kann, sondern die Populationswerte der Standardfehler anhand der Replikationen geschätzt werden müssen, wodurch eine gewisse Ungenauigkeit des zugrundeliegenden Vergleichswerts immer gegeben ist. Aufgrund der vergleichsweise geringen Replikationszahl von 300 erscheint es auch noch angemessen, relative Biasraten, welche leicht über 10 % liegen, noch als akzeptabel anzusehen.

Tabelle 30: Relativer SE-Bias nach Invarianz-Stufen

Invar	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(I_C3)	M(S_C1)	M(S_C2)	M(S_C3)	N
1.000	0.008	-0.018	0.078	-0.026	-0.022	-0.007	0.019	-0.098	0.106	300
2.000	0.006	-0.016	0.079	-0.023	-0.018	-0.005	0.029	-0.101	0.105	300
3.000	0.007	-0.011	0.074	-0.026	-0.022	-0.004	0.036	-0.096	0.087	300
4.000	0.016	-0.025	-0.025	-0.356	-0.380	-0.011	-0.850	-0.864	0.116	240
5.000	-0.082	-0.027	-0.153	-0.365	-0.153	-0.047	-0.475	-0.623	0.092	152
6.000	-0.105	-0.031	-0.021	-0.388	-0.542	-0.359	-0.442	-0.900	-0.368	144

Info: Klassendistanz=4 MD, Mixing=33.33.33, Nobs=1300.

Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

Invarianzstufen. Bei hoher Klassendistanz, 1300 simulierten Beobachtungsfällen und einer prozentualen Gleichverteilung dieser Fälle über die latenten Klassen scheinen die Standardfehler der Varianzparameter für keine der simulierten Invarianzstufen deutlich verzerrt geschätzt zu werden. Lediglich für die kombinierte Invarianzstufe 6 wird der Grenzwert für die Standardfehlerschätzung der Interceptvarianz knapp erreicht (vgl. Tabelle 30 auf Seite 154).

Für die längsschnittlichen Simulationskonfigurationen zeigt sich insgesamt, dass auch die Standardfehler der Mittelwerte für die Intercepts und die Slopes nicht stark verzerrt geschätzt werden. Zwar wird der Grenzwert für einzelne Slopes knapp erreicht, aber dies stellt keinen deutlichen Unterschied zur unverzerrten Invarianzstufe 1 mit Messgleichheit dar.

Für die gruppenspezifischen Invarianzformen zeigen sich hingegen deutlich verzerrte Standardfehlerschätzungen für die Mittelwerte der Intercepts und Slopes. Die Standardfehler der manipulierten Klassen werden im Schnitt deutlich zu klein geschätzt. Dies gilt sowohl für die Standardfehler der Intercepts als auch für die Standardfehler der Slopes, wobei der Grad der Verzerrung für die Slopes noch etwas größer ausfällt. Während die Klassen (Mixing=1) und große Klassendistanz von 4.0 MD (Sep=3).

Standardfehler der Intercepts und Slopes der nicht direkt manipulierten latenten Klasse (Referenzklasse 3) für die gruppenspezifischen Nicht-Invarianzen noch nicht stark verzerrt geschätzt werden, wandelt sich dies für die kombinierte Invarianzkonfiguration. Das Maß der Verzerrung gleicht in der 6. Invarianzstufe dem Grad der Verzerrung in den direkt manipulierten Klassen.

Zusammenfassend kann festgehalten werden, dass die Invarianzformen insgesamt keinen nennenswerten Einfluss auf die Standardfehler der Varianzterme nehmen. Die längsschnittlichen Invarianzformen scheinen zudem die Standardfehler der Intercepts und Slopes nicht nennenswert zu verzerren. Daher sind die Signifikanztests für diese Ausprägungen als zuverlässig einzustufen. Für die gruppenspezifischen Invarianzformen und die kombinierte Form zeigen sich hingegen deutliche Probleme bzgl. zu kleiner Standardfehlerschätzungen der Intercept- und Slopeparameter. Der Typ 1 Fehler wird hierbei deutlich erhöht, wodurch keine zuverlässigen Inferenzschlüsse mehr möglich sind.

Tabelle 31: Relativer SE-Bias nach Klassendistanz-Stufen

Sep	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(I_C3)	M(S_C1)	M(S_C2)	M(S_C3)	N
1.000	-0.408	-0.486	-0.546	-0.716	-0.596	-0.611	-0.679	-0.868	-0.763	62
2.000	-0.234	-0.360	-0.329	-0.954	-0.547	-0.603	-0.895	-0.783	-0.787	91
3.000	0.008	-0.018	0.078	-0.026	-0.022	-0.007	0.019	-0.098	0.106	300

Info: Invarianz=1, Mixing=33.33.33, Nobs=1300.

Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

Klassendistanz. Für die verschiedenen Ausprägungen der Klassendistanz zeigen sich deutliche Unterschiede bzgl. der Verzerrung der Standardfehlerschätzungen (vgl. Tabelle 31 auf Seite 155). Für die größte Klassendistanz von 4.0 MD werden die Standardfehler für keinen der Wachstumsparameter verzerrt geschätzt. Auch wenn sich die Standardfehler der Slopemittelwerte dem Grenzwert tlw. annähern, kann von keinen substantiell bedeutenden Verzerrungen gesprochen werden.

Gänzlich anders sieht dies für die Klassendistanzstufen 0.8 und 1.5 MD aus. Die Standardfehler aller Wachstumsparameter werden im Schnitt deutlich zu klein geschätzt, was zu einer Erhöhung des Fehlers erster Art führt und die Nullhypothese im Schnitt zu oft verworfen wird. Während das Ausmaß dieser Verzerrung für die Standardfehler der Varianzparameter mit kleinerer Klassendistanz zunimmt, zeigt sich für die Standardfehler der Mittelwertschätzungen der Intercepts und Slopes kein systematischer Unterschied zwischen der mittleren und der niedrigsten Klassendistanz. Es bleibt festzuhalten, dass

die Signifikanztests für große Klassendistanzen unverzerrt und gültig bleiben. Bereits für die mittlere Klassendistanz von 1.5 MD können keine zuverlässigen Signifikanztests mehr durchgeführt werden, selbst wenn gänzliche Messinvarianz, hohe Fallzahl und gleichverteilte Klassenanteile gegeben sind. Einschränkend muss allerdings erwähnt werden, dass die durchschnittlichen prozentualen Verzerrungen für die mittlere und niedrigste Klassendistanzstufe lediglich auf 91 und 62 Replikationen beruhen und damit ein höheres Maß an Unsicherheit mit den Durchschnittswerten verbunden ist.

Tabelle 32: Relativer SE-Bias nach Fallzahl-Stufen

Nobs	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(I_C3)	M(S_C1)	M(S_C2)	M(S_C3)	N
1.000	-0.014	0.072	0.061	-0.070	0.055	0.043	0.032	0.015	0.108	297
2.000	0.093	-0.064	0.026	-0.080	0.018	-0.011	-0.040	-0.087	0.145	299
3.000	0.008	-0.018	0.078	-0.026	-0.022	-0.007	0.019	-0.098	0.106	300

Info: Invarianz=1, Mixing=33.33.33, Klassendistanz=4 MD.

Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

Fallzahl und Klassenanteilsverhältnisse. Unter ansonsten weitestgehend günstigen Simulationsbedingungen scheinen das Klassenanteilsverhältnis sowie die Fallzahl kaum einen nennenswerten verzerrenden Einfluss auf die Standardfehlerschätzungen der Varianz-, Intercept- und Slopeparameter zu haben (vgl. hierzu Tabelle 32 und 33). Der Grenzwert des relativen SE-Bias wird insbesondere für die Varianzen und die Interceptmittelwerte deutlich unterschritten. Auch die Standardfehler der Slopemittelwerte bleiben, zum Großteil, unter der veranschlagten Interpretationsschwelle. Lediglich Standardfehler der Slopemittelwerte der Referenzklasse 3 überschreiten den Grenzwert leicht, ohne dabei jedoch einen besonders starken Effekt oder eine klare Systematik aufzuweisen. Insgesamt kann festgehalten werden, dass weder Fallzahl noch Klassenanteilsverhältnisse die Schätzungen der Standardfehler der Wachstumsverlaufparameter unter ansonsten günstigen Bedingungen wesentlich beeinflussen und die Signifikanztests somit unverzerrt sind.

Tabelle 33: Relativer SE-Bias nach Mixing-Stufen

Mixing	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(I_C3)	M(S_C1)	M(S_C2)	M(S_C3)	N
1.000	0.008	-0.018	0.078	-0.026	-0.022	-0.007	0.019	-0.098	0.106	300
2.000	0.007	-0.009	0.063	-0.040	-0.019	-0.039	0.036	-0.020	0.037	300
3.000	0.032	-0.005	0.041	-0.065	-0.030	-0.021	0.101	0.071	0.070	299

Info: Invarianz=1, Nobs=1300, Klassendistanz=4 MD.

Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

11.2 Modellbasiert Analyse bzgl. des SE-Bias

Auch für die Verzerrungen der Standardfehler wird eine zu den deskriptiven Ergebnissen ergänzende modellbasierte Auswertung durchgeführt, welche alle Simulationskonfigurationen enthält und als Auswertungsbasis die einzelnen Replikationen umfasst. Die Zieldimension ist die absolute SE-Verzerrung, welche aus Gründen der einfacheren Interpretation als absoluter Betrag berücksichtigt wird. Damit sind negative Regressionskoeffizienten ein Ausdruck von weniger verzerrten Standardfehlern und vice versa. Die unabhängigen Modellvariablen setzen sich jeweils aus allen Haupteffekten der Simulationsgrößen sowie den substantiell bedeutsamen Interaktionstermen zwischen den Simulationsgrößen zusammen. Interaktionsterme werden nur dann berücksichtigt, wenn sie mindestens 2 % (ΔR^2) mehr Varianzaufklärung zu den Haupteffekten beitragen, was noch als kleine Effektstärke eingeordnet werden kann (vgl. Cohen, 1988: 413 ff.). Die Ergebnisse der folgenden Metamodelle sind mit unstandardisierten Koeffizienten in den Tabellen 34, 37 und 39 ausgewiesen. Die relative Bedeutsamkeit der Effekte der Metamodelle ist in den Tabellen 35, 36 und 38 mit den ΔR^2 -Werten dargestellt.⁷⁸

11.2.1 Analyse bzgl. des SE-Bias der Ko-/Varianzparameter

Die Varianz der Verzerrung der Standardfehler der Intercept- und Slopevarianzterme wird nur moderat durch die Simulationsgrößen statistisch aufgeklärt, was sich an den R^2 -Werten der Metamodelle zwischen 16 und 18 % festmachen lässt (vgl. Tabelle 34 auf Seite 158). Die Standardfehler der Kovarianz zwischen Intercept und Slope wird hingegen mit einem Wert von knapp 52 % deutlich stärker durch die Simulationsgrößen bestimmt.

Über alle drei Modelle hinweg zeigt sich, dass eine höhere Fallzahl im Schnitt mit weniger verzerrten Standardfehlern einhergeht, was sich in der deskriptiven Auswertung (unter optimalen Bedingungen) des relativen SE-Bias noch nicht zeigte (vgl. Tabelle 32). Kleine und ungleiche Klassenanteile haben hingegen keinen nennenswerten Effekt auf die Standardfehler der Varianzterme, was bereits in den deskriptiven Auswertungen deutlich geworden ist. Die zusätzliche Varianzaufklärung (ΔR^2) der Klassenanteile übersteigt für keines der Modelle 0,3 % (vgl. Tabelle 35 auf Seite 159). Auch eine höhere Klassendistanz ist mit geringeren Verzerrungen der Standardfehler der Varianzterme assoziiert. Die Stärke dieses Einflusses übersteigt in allen drei Modellen den Einfluss der Fallzahl.

⁷⁸Eine visuelle Diagnostik zu allen Meta-Modellen kann aus dem Online-Appendix A.5 entnommen werden.

Tabelle 34: Regressions-Ergebnisse zu SE-Bias der Ko-/Varianz Parameter (robuste SE; HC3)

	<i>Dependent variable:</i>		
	model_SE_i_v	model_SE_s_v	model_SE_is_cv
	(1)	(2)	(3)
Constant	0.039*** (0.001)	0.010*** (0.0001)	0.015*** (0.0003)
nobs2	-0.005*** (0.0005)	-0.001*** (0.0001)	-0.002*** (0.0001)
nobs3	-0.010*** (0.0002)	-0.003*** (0.00004)	-0.003*** (0.0001)
mixing2	-0.003*** (0.0005)	-0.0004** (0.0001)	-0.001*** (0.0001)
mixing3	-0.003*** (0.0005)	-0.0002 (0.0001)	-0.001*** (0.0001)
invar2	0.0003 (0.001)	0.0003 (0.0002)	0.001 (0.0005)
invar3	0.005*** (0.001)	-0.001** (0.0002)	0.0002 (0.0004)
invar4	0.001 (0.001)	0.0003 (0.0002)	-0.001 (0.0004)
invar5	-0.014*** (0.001)	-0.005*** (0.0002)	-0.005*** (0.0004)
invar6	-0.012*** (0.001)	-0.005*** (0.0002)	-0.007*** (0.0004)
sep2	-0.019*** (0.001)	-0.003*** (0.0002)	-0.006*** (0.0004)
sep3	-0.030*** (0.001)	-0.008*** (0.0001)	-0.012*** (0.0004)
invar2:sep2	-0.001 (0.001)	-0.001*** (0.0003)	-0.001 (0.001)
invar3:sep2	-0.002 (0.001)	0.0001 (0.0003)	0.001* (0.001)
invar4:sep2	-0.003*** (0.001)	-0.001*** (0.0003)	-0.0001 (0.0005)
invar5:sep2	0.025*** (0.003)	0.008*** (0.001)	0.011*** (0.001)
invar6:sep2	0.025*** (0.001)	0.005*** (0.0002)	0.012*** (0.0005)
invar2:sep3	-0.0003 (0.001)	-0.0002 (0.0002)	-0.001 (0.0005)
invar3:sep3	-0.004*** (0.001)	0.002*** (0.0002)	0.0001 (0.0004)
invar4:sep3	-0.001 (0.001)	-0.0003 (0.0002)	0.001 (0.0004)
invar5:sep3	0.019*** (0.001)	0.005*** (0.0002)	0.008*** (0.0004)
invar6:sep3	0.018*** (0.001)	0.006*** (0.0002)	0.009*** (0.0004)
Observations	23,084	23,084	23,084
R ²	0.166	0.185	0.517
Adjusted R ²	0.165	0.184	0.516
Residual Std. Error (df = 23062)	0.025	0.007	0.005
F Statistic (df = 21; 23062)	218.038***	248.579***	1,175.067***

Info:

*p<0.05; **p<0.01; ***p<0.001

Aus der gemeinsamen Betrachtung des Haupteffekts der Invarianzformen und dem Interaktionseffekt mit der Klassendistanz wird klar, dass die längsschnittlichen Invarianzformen keinen nennenswerten Einfluss auf die Standardfehlerschätzung der drei Varianzterme haben, auch unabhängig von der Klassendistanz. Gleiches gilt auch für die gruppenspezifische und kombinierte Invarianzform für hohe Klassendistanzen. Für geringere Klassendistanzen zeigt sich allerdings, dass gruppenspezifische und kombinierte Messunterschiede sogar einen positiven Einfluss haben. Die Standardfehler werden für dieses Simulationsausprägungen weniger verzerrt geschätzt. Die Messunterschiede erzeugen für geringe Klassendistanzen vermutlich zusätzliche Klassendifferenz, welche eine realitätsnähere Heterogenitätsschätzung der Varianzen begünstigt, wodurch die Standardfehlerschätzung besser ausfällt.

Insgesamt zeigt sich, dass die Simulationsgrößen die Standardfehlerschätzungen der Varianzen nur mäßig beeinflussen, was auch schon in der deskriptiven Auswertung des relativen SE-Fehlers deutlich wurde. Die Schätzung über alle Konfigurationen hinweg macht allerdings auch deutlich, dass höhere Klassendistanz und höhere Fallzahl im Schnitt zu weniger verzerrten Standardfehlern der Varianzterme führt, was in den unifaktoriellen deskriptiven Auswertungen so noch nicht deutlich wurde. Klassenanteile haben auch anhand der Metamodelle keinen bedeutenden Einfluss. Der Effekt von Messunterschieden zeigt sich nur bei geringer Klassendistanz und nur, wenn Messunterschiede zwischen den Gruppen vorhanden sind. Daraus lässt sich schlussfolgern, dass eine angemessene Fallzahl und höhere Klassendistanz zu zuverlässigen Standardfehlerschätzungen führen. Selbst Messunterschiede haben entweder keinen oder wenn, dann einen positiven Einfluss auf die Standardfehlerschätzungen.

Tabelle 35: Delta-R-Quadrat der Praediktoren zum SE-Bias der Ko-/Varianz Parameter

	df	dR Var I	dR Var S	dR KoVar IS
nobs	2	0.019	0.021	0.043
mixing	2	0.003	0.0005	0.003
invar	5	0.014	0.022	0.044
sep	2	0.022	0.027	0.067
invar:sep	10	0.028	0.033	0.090

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 34.

11.2.2 Analyse bzgl. des SE-Bias der Interceptfaktor-Mittelwerte

Im Vergleich zu den Verzerrungen der Standardfehler der Varianzterme zeigt sich für die Modelle zu den Standardfehlern der Interceptmittelwerte eine deutlich höhere Varianzaufklärung durch die Simulationsgrößen. Die R^2 -Werte bewegen sich im Bereich von 43 bis 60 % für die Standardfehlerunterschiede der Interceptterme der drei latenten Klassen (vgl. Tabelle 37 auf Seite 162).

Die simulierte Fallzahl hat für keine der latenten Klassen einen nennenswerten (und damit interpretationswürdigen) Einfluss auf die Verzerrung der Standardfehler der Interceptterme, obwohl die Tendenz anzeigt, dass höhere Fallzahlen eher mit geringeren Verzerrungen verbunden sind. Große Klassendistanzen sind hingegen durchgehend mit geringeren Verzerrungen verbunden. Die mit der Klassendistanz verbundenen Effekte sind für alle drei Modelle als vergleichsweise bedeutsam einzustufen, was sich an den entsprechenden Delta- R^2 -Werten festmachen lässt (vgl. Tabelle 36 auf Seite 161). Kleine Klassenanteile zeigen wiederum nur für Standardfehler der ersten beiden Klassen einen nennenswerten Effekt, wohingegen die dritte Referenzklasse davon nicht betroffen ist. So haben die Haupteffekte der Klassenanteile für die Klassen 1 und 2 zwar keinen eigenen bedeutenden Effekt auf die Standardfehlerverzerrung, aber die Interaktion zwischen der Klassendistanz und den Klassenanteilen macht deutlich, dass sich für große Klassendistanzen kleine Klassenanteile negativ auf die Schätzungen der Standardfehler auswirken. Die zusätzliche Varianzaufklärung dieser Interaktion beträgt für beide Modelle ca. 4 %.

Für die Messinvarianzstufen und deren Interaktion mit der Klassendistanz zeigt sich im Durchschnitt kein einheitliches Muster. Die Verzerrungen der Standardfehler werden für die einzelnen Ausprägungen sowohl verstärkt wie auch verringert. Aus einer zusammenfassenden Betrachtung wird lediglich deutlich, dass für große Klassendistanzen eher größere Verzerrungen vorherrschen. Darüber hinaus werden die Verzerrungen der Standardfehler der nicht direkt manipulierten latenten Referenzklasse 3 weniger stark durch die Invarianzformen beeinflusst. Auch wenn sich kein klares Muster ausmachen lässt, können die Invarianzformen dennoch einen bedeutsamen Anteil an Varianz der Verzerrungen binden. Daher darf das Fehlen von eindeutigen Mustern nicht als fehlende Bedeutung der Simulationsgröße interpretiert werden.

Zusammenfassend lässt sich festhalten, dass vor allem große Klassendistanzen zu zuverlässigen SE-Schätzungen führen, wobei allerdings kleine Klassenanteile ein Problem werden können, da sie SE-Schätzungen verzerren. Die Fallzahl scheint auch nach der

modellbasierten Auswertungen keinen nennenswerten Einfluss auf Verzerrungen der Standardfehler der Interceptterme zu nehmen, was dafür spricht, dass zuverlässige Signifikanztests der Mittelwerte der Intercepts auch bei einer Fallzahl von lediglich 300 durchführbar sind. Fehlende Invarianz der Messungen über die Zeit, zwischen den Gruppen und in der Kombination hat in jedem Fall einen Einfluss auf die SE-Schätzungen, weist aber keine einheitliche Tendenz auf. Folglich sollte nach Möglichkeit Invarianz der Messung angestrebt werden, um zuverlässige Signifikanztests für die Mittelwerte der Intercepts sicher durchführen zu können.

Tabelle 36: Delta-R-Quadrat der Praediktoren zum SE-Bias der Intercept Parameter

	df	dR Mean I c1	dR Mean I c2	dR Mean I c3
nobs	2	0.005	0.013	0.019
mixing	2	0.001	0.0004	0.012
invar	5	0.003	0.057	0.016
sep	2	0.094	0.036	0.093
invar:sep	10	0.149	0.022	0.034
mixing:sep	4	0.042	0.037	

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 37.

Tabelle 37: Regressions-Ergebnisse zum SE-Bias der Interceptfaktoren (robuste SE; HC3)

	<i>Dependent variable:</i>		
	model_SE_i_m_c1	model_SE_i_m_c2	model_SE_i_m_c3
	(1)	(2)	(3)
Constant	0.427*** (0.009)	0.129*** (0.003)	0.282*** (0.004)
nobs2	-0.013* (0.007)	-0.018*** (0.001)	-0.032*** (0.002)
nobs3	-0.084*** (0.007)	-0.028*** (0.001)	-0.049*** (0.001)
mixing2	-0.004 (0.009)	0.009** (0.003)	0.001 (0.002)
mixing3	-0.067*** (0.008)	0.010* (0.004)	0.035*** (0.002)
invar2	0.002 (0.009)	0.002 (0.003)	0.002 (0.006)
invar3	0.077*** (0.009)	0.017*** (0.003)	0.046*** (0.006)
invar4	0.139*** (0.013)	0.008* (0.004)	-0.004 (0.006)
invar5	-0.052*** (0.010)	0.089*** (0.003)	-0.074*** (0.005)
invar6	0.081*** (0.012)	0.131*** (0.006)	-0.053*** (0.006)
sep2	0.787*** (0.049)	-0.001 (0.004)	-0.018** (0.006)
sep3	-0.462*** (0.009)	-0.118*** (0.003)	-0.257*** (0.004)
invar2:sep2	-0.108 (0.057)	-0.001 (0.004)	-0.031*** (0.008)
invar3:sep2	-0.709*** (0.051)	-0.012** (0.004)	-0.091*** (0.008)
invar4:sep2	-0.852*** (0.049)	-0.065*** (0.004)	-0.033*** (0.008)
invar5:sep2	-0.728*** (0.047)	-0.078*** (0.005)	0.042*** (0.009)
invar6:sep2	-0.800*** (0.048)	-0.042*** (0.007)	0.074*** (0.008)
invar2:sep3	0.0005 (0.009)	0.003 (0.003)	-0.007 (0.006)
invar3:sep3	-0.009 (0.010)	0.029*** (0.003)	-0.045*** (0.006)
invar4:sep3	-0.135*** (0.013)	-0.007 (0.004)	0.005 (0.006)
invar5:sep3	0.609*** (0.017)	-0.062*** (0.003)	0.127*** (0.005)
invar6:sep3	0.416*** (0.014)	-0.043*** (0.006)	0.127*** (0.006)
mixing2:sep2	-0.220*** (0.018)	-0.012** (0.004)	
mixing3:sep2	0.485*** (0.025)	-0.046*** (0.005)	
mixing2:sep3	0.042*** (0.009)	-0.006 (0.003)	
mixing3:sep3	0.302*** (0.010)	0.052*** (0.004)	
Observations	23,084	23,084	23,084
R ²	0.434	0.542	0.602
Adjusted R ²	0.433	0.542	0.602
Residual Std. Error	0.390 (df = 23058)	0.067 (df = 23058)	0.091 (df = 23062)
F Statistic	707.472*** (df = 25; 23058)	1,092.264*** (df = 25; 23058)	1,661.663*** (df = 21; 23062)
Info:	*p<0.05; **p<0.01; ***p<0.001		

11.2.3 Analyse bzgl. des SE-Bias der Slopefaktor-Mittelwerte

Auch für die Verzerrungen der Standardfehler der Slopemittelwerte bewegt sich die Varianzaufklärung im Bereich von 40 bis 60 % und kann somit insgesamt ähnlich gut statistisch durch die Simulationsfaktoren erklärt werden, wie die Standardfehlerverzerrung der Interceptmittelwerte (vgl. Tabelle 39 auf Seite 165). Wie die deskriptive Darstellung zum relativen SE-Bias schon angedeutet hat, zeigt die Simulationsgröße «Fallzahl» für die Standardfehlerschätzungen der Slopemittelwerte in keiner latenten Klasse einen bedeutenden Einfluss. Die zusätzliche Varianzaufklärung bleibt jeweils deutlich unter der Interpretationsschwelle.

Für die Klassenanteilsverhältnisse besteht hingegen keine klare Übereinstimmung mit den deskriptiven Tabellenanalysen. Während der relative SE-Bias für die optimalen Simulationsbedingungen noch keinen bedeutenden Effekt erkennen ließ, gestaltet sich das Bild auf Grundlage der Metamodelle etwas differenzierter. Für die dritte latente Referenzklasse lässt sich analog zu der deskriptiven Auswertung durchschnittlich zwar kein relevanter Einfluss von den Klassenanteilsverhältnissen ausmachen, aber für die erste und die zweite latente Klasse lassen sich interpretationswürdige Einflüsse identifizieren. Dabei sticht nicht so sehr der Haupteffekt der Klassenanteilsverhältnisse hervor, sondern die Interaktion mit der Klassendistanz und den Invarianzformen. So zeigt sich anhand der Metamodelle, dass insbesondere für hohe Klassendistanzen geringe Klassenanteile einen verzerrenden Einfluss auf die Standardfehlerschätzungen haben. Für die erste latente Klasse zeigt sich zudem, dass die Invarianzformen insbesondere dann verzerrend wirken, wenn sie mit kleinen Klassenanteilen im Verbund auftreten. Die kombinierte zusätzliche Varianzaufklärung beträgt für das Modell der ersten latenten Klasse immerhin 24 %, wohingegen die zusätzliche Varianzaufklärung für das Metamodell der zweiten latenten Klasse zumindest mit ca. 2 % zu Buche schlägt (vgl. Tabelle 38 auf Seite 164). Dies ist insofern beachtenswert, als dass kleine Klassenanteile hinsichtlich der kumulierten Varianzaufklärung nicht in der direkt manipulierten kleinen (latenten zweiten) Klasse den stärksten Effekt aufweist, sondern vor allem in der ersten Klasse. Die Standardfehler der dritten Referenzklasse werden hingegen nicht von der Simulationsvariable «Anteilsverhältnisse» beeinflusst.

Den größten Anteil an Varianz der Metamodelle bindet allerdings wiederum die Klassendistanz im Verbund mit der Interaktion zwischen der Klassendistanz und den Invarianzstufen. Hohe Klassendistanz ist in allen drei Analysemodellen mit geringeren Verzerrungen der Standardfehler der Slopemittelwerte verbunden. Die Koeffizienten der Interaktionen

zwischen der Klassendistanz und den Invarianzstufen zeigen, dass deutliche Unterschiede zwischen den Effekten für geringe und hohe Klassendistanzen existieren, unabhängig davon, welche latente Klasse betrachtet wird. Für niedrige Klassendistanzen werden die Standardfehler bei fehlender Messinvarianz im Schnitt für alle Klassen sogar weniger verzerrt geschätzt. Dieses Ergebnis dreht sich allerdings bei hoher Klassendistanz. Für die direkt manipulierten Klassen 1 und 2 zeigen insbesondere (aber nicht ausschließlich) für die fehlenden Gruppeninvarianzausprägungen stärker verzerrte Standardfehler der Slopemittelwerte. Für die dritte Referenzklasse verändert sich der Effekt der Invarianzstufen auch in Abhängigkeit von der Klassendistanz, aber für die höchste Klassendistanz sind keine stärker verzerrten Standardfehler auszumachen. Vielmehr zeigt sich bei hoher Klassendistanz, dass die Standardfehler der dritten Klasse kaum durch die Invarianzstufen beeinflusst werden.

Insgesamt lässt sich festhalten, dass die Standardfehler der Slopemittelwerte zu einem nicht zu vernachlässigbaren Anteil von den Simulationsgrößen beeinflusst werden. Mit der Ausnahme der Größe «Fallzahl» haben alle Simulationsgrößen daran einen Anteil. Der größte Anteil der Varianzaufklärung ist dabei auf die Klassendistanz, die Invarianzstufen und die Interaktion zwischen beiden zurückzuführen. Für die ersten beiden latenten Klassen spielen auch kleine Klassengrößen eine Rolle für die Verzerrungen der Standardfehler der Slopemittelwerte. Dabei wirken sich kleine Klassenanteile besonders dann verzerrend aus, wenn die Klassendistanzen groß sind.

Tabelle 38: Delta-R-Quadrat der Praediktoren zum SE-Bias der Slope Parameter

	df	dR Mean S c1	dR Mean S c2	dR Mean S c3
nobs	2	0.001	0.015	0.004
mixing	2	0.028	0.002	0.012
invar	5	0.030	0.016	0.080
sep	2	0.037	0.062	0.174
invar:sep	10	0.163	0.178	0.077
mixing:sep	4	0.088	0.019	
mixing:invar	10	0.158		

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 39.

Tabelle 39: Regressions-Ergebnisse zum SE-Bias der Slopefaktoren (robuste SE; HC3)

	<i>Dependent variable:</i>		
	model_SE_s_m_c1	model_SE_s_m_c2	model_SE_s_m_c3
	(1)	(2)	(3)
Constant	0.607*** (0.019)	0.362*** (0.016)	0.701*** (0.039)
nobs2	−0.034*** (0.006)	−0.018*** (0.003)	−0.028*** (0.003)
nobs3	−0.024*** (0.007)	−0.067*** (0.002)	−0.034*** (0.003)
mixing2	−0.208*** (0.014)	−0.061*** (0.008)	0.012*** (0.002)
mixing3	−0.830*** (0.020)	−0.025** (0.008)	0.061*** (0.003)
invar2	0.024 (0.023)	−0.002 (0.017)	−0.453*** (0.041)
invar3	−0.208*** (0.022)	−0.062*** (0.013)	−0.318*** (0.043)
invar4	−0.139*** (0.021)	0.056*** (0.016)	−0.304*** (0.044)
invar5	−0.596*** (0.021)	−0.139*** (0.013)	−0.490*** (0.041)
invar6	−0.712*** (0.023)	−0.110*** (0.013)	−0.459*** (0.041)
sep2	0.142*** (0.025)	−0.008 (0.022)	−0.348*** (0.045)
sep3	−0.658*** (0.018)	−0.369*** (0.017)	−0.689*** (0.041)
invar2:sep2	−0.001 (0.032)	0.054 (0.031)	0.316*** (0.046)
invar3:sep2	0.390*** (0.037)	0.067** (0.024)	0.132** (0.048)
invar4:sep2	−0.166*** (0.028)	−0.145*** (0.023)	0.216*** (0.049)
invar5:sep2	−0.086** (0.027)	−0.038 (0.021)	0.364*** (0.046)
invar6:sep2	0.028 (0.029)	−0.061** (0.021)	0.360*** (0.046)
invar2:sep3	−0.025 (0.021)	0.007 (0.018)	0.456*** (0.041)
invar3:sep3	0.476*** (0.019)	0.137*** (0.013)	0.349*** (0.043)
invar4:sep3	0.199*** (0.020)	0.163*** (0.016)	0.301*** (0.044)
invar5:sep3	1.040*** (0.027)	0.447*** (0.014)	0.502*** (0.041)
invar6:sep3	1.363*** (0.032)	0.455*** (0.014)	0.528*** (0.041)
mixing2:sep2	−0.159*** (0.015)	0.064*** (0.009)	
mixing3:sep2	0.483*** (0.020)	0.188*** (0.014)	
mixing2:sep3	0.223*** (0.014)	0.076*** (0.008)	
mixing3:sep3	1.075*** (0.018)	0.204*** (0.008)	
mixing2:invar2	−0.030* (0.012)		
mixing3:invar2	0.041* (0.017)		
mixing2:invar3	−0.157*** (0.016)		
mixing3:invar3	0.796*** (0.024)		
mixing2:invar4	0.150*** (0.010)		
mixing3:invar4	0.272*** (0.017)		
mixing2:invar5	0.454*** (0.023)		
mixing3:invar5	1.105*** (0.026)		
mixing2:invar6	0.304*** (0.019)		
mixing3:invar6	1.561*** (0.032)		
Observations	23,084	23,084	23,084
R ²	0.606	0.399	0.409
Adjusted R ²	0.605	0.399	0.408
Residual Std. Error	0.379 (df = 23048)	0.180 (df = 23058)	0.180 (df = 23062)
F Statistic	1,012.638*** (df = 35; 23048)	613.056*** (df = 25; 23058)	759.832*** (df = 21; 23062)
Info:	*p<0.05; **p<0.01; ***p<0.001		

11.3 Diskussion zu den modellbasierten Analysen bzgl. des SE-Bias

Wie bereits für die Auswertung des Bias sind die Ergebnisse zur SE-Bias vor dem Hintergrund der korrekten Klassenidentifikation zu interpretieren (eine ausführlichere Diskussion findet sich in Kapitel 10.3). Im Überblick zeigt sich, dass eine getrennte Betrachtung der latenten Klassen auch im Kontext des SE-Bias angebracht ist. Sowohl die Effekte der Simulationsgrößen als auch der Anteil an statistisch erklärter Varianz hinsichtlich der SE-Schätzungen unterscheidet sich tlw. zwischen den drei latenten Klassen. Dennoch lassen sich in einer zusammenfassenden Betrachtung der durchgeführten Analysen auch einige Gemeinsamkeiten ausmachen. Während die Verzerrungen der Standardfehler der Intercept- und Slopevarianzen sich mit lediglich 16 bis 18 % Varianzaufklärung vergleichsweise robust gegenüber den Simulationsgrößen zeigen, sind die Verzerrungen der Standardfehler der Intercept- und Slopemittelwerte mit 40 bis 60 % Varianzaufklärung deutlich sensibler gegenüber der Variation der Simulationsgrößen. Im Folgenden werden die Unterschiede und Gemeinsamkeiten hinsichtlich der Effekte der Simulationsgrößen im Überblick dargestellt. Die wesentlichen Aspekte dieser Zusammenfassung sind auch in Tabelle 40 aufgeführt.

Die Simulationsgröße «Fallzahl» ist für die SE-Schätzungen der Intercept- und Slopemittelwerte weitestgehend unbedeutend, wohingegen die geschätzten Standardfehler der Kovarianz und der Varianzen der Intercepts und Slopes positiv auf eine höhere Fallzahl reagieren. Je höher die Fallzahl der Simulation, desto unverzerrter werden die Standardfehler im Mittel geschätzt.

Für alle latenten Klassen und alle Standardfehlerschätzungen der betrachteten Zielparameterschätzungen führt eine große Klassendistanz zu weniger verzerrten Schätzungen. Die eigenständige Varianzaufklärung der Klassendistanz weist diese Simulationsgröße über alle Metamodelle hinweg entweder als stärkste oder als eine der stärksten Einflussgrößen aus. Wie bereits für die Klassenidentifikation oder die Parameterverzerrungen stellt eine große Klassendistanz eine essenzielle Voraussetzung für eine zuverlässige Schätzung der Standardfehler der Wachstumsfaktoren von MIGMMs dar.

Kleine Klassenanteile haben hinsichtlich der Standardfehlerverzerrungen der Varianzterme keine nennenswerte Bedeutung. Für die Standardfehler der Mittelwertschätzung von Intercept und Slope lassen sich allerdings für die ersten beiden latenten Klassen relevan-

te Einflüsse feststellen. Für Simulationskonfigurationen mit großer Klassendistanz zeigt sich für die Standardfehlerverzerrungen der Intercept- und Slopemittelwerte, dass sie bei kleinen Klassen (5 %) deutlich verzerrter geschätzt werden. Die Simulation einer kleinen Klassengröße wirkt sich dabei folglich nicht nur in der als klein simulierten Klasse auf die Schätzung der Standardfehler aus, sondern strahlt im Durchschnitt auf die erste Nachbarklasse aus.

Wie bereits bei den Analysen zum Parameterbias variiert der Einfluss der Invarianzformen in Abhängigkeit von der Klassendistanz. Des Weiteren unterscheidet sich der Einfluss zwischen den drei betrachteten latenten Klassen und der Art der den Standardfehlern zugrundeliegenden Parameterschätzungen (Varianzen vs. Mittelwerte). Zudem ist die relative Bedeutsamkeit dieses Effekts über die neun Metamodelle hinweg stets auf den vorderen Plätzen zu verorten. Für die Standardfehler der Varianzterme zeigt sich, dass die Invarianzformen bei hohen Klassendistanzen im Schnitt keinen relevanten Effekt auf die Verzerrung haben. Diese Schlussfolgerung lässt sich auch aus den deskriptiven Auswertungen des relativen SE-Bias unter optimalen Simulationsbedingungen ableiten. Unter der Voraussetzung von geringer Klassendistanz kann fehlende Messinvarianz sogar eine verzerrungsfreiere Schätzung der Standardfehler begünstigen, solange Messunterschiede zwischen den latenten Gruppen bestehen. Folglich kann sich fehlende Messinvarianz lediglich positiv (oder zumindest neutral) auf den Inferenzschluss zur Heterogenität der Wachstumsfaktoren auswirken. Anders sieht es hingegen für die Standardfehler der Intercept- und Slopemittelwerte aus. Hier deutet sich an, dass insbesondere für hohe Klassendistanzen die Standardfehler für die nicht-invarianten Messkonfigurationen verzerrt geschätzt werden. Dies trifft allerdings nicht auf die dritte Referenzklasse zu, welche für die vierte bis sechste Messinvarianzstufe nicht direkt manipuliert wurde. Bei hoher Klassendistanz ist die Verzerrung durch fehlende Messinvarianz für die dritte Klasse eher als vernachlässigbar einzustufen. Lässt man die Simulationskonfigurationen mit hoher Klassendistanz beiseite und betrachtet nur die geringeren Distanzen, so zeigt sich für die Standardfehler der Interceptsmittelwerte ein uneinheitliches Muster von Unter- und Überschätzung und für die Slopemittelwerte eine deutliche Tendenz zur verzerrungsfreieren Schätzung der Standardfehler (auch unabhängig von der betrachteten latenten Klasse).

Resümierend lässt sich festhalten, dass insbesondere die Standardfehler der Mittelwerte der Wachstumsfaktoren durch die Simulationsgrößen beeinflusst werden. Ein zusammen-

fassender Überblick über die hier diskutierte Performanz bzgl. des SE-Bias der Wachstumsfaktoren kann auch aus Tabelle 40 entnommen werden.

Tabelle 40: Zusammenfassung der SE-Bias-Performanz bzgl. der Wachstumsfaktoren

	Kovarianz/Varianzen	Intercept-Mittelwerte	Slope-Mittelwerte
Fehlende Messinvarianz	zeitlich	unbedeutend	uneinheitlich (für hohe Sep verzerrter) verzerrter (bei hoher Sep)
	gruppenspez.	unbedeutend (bei geringer Sep: unverzerrter)	uneinheitlich (für hohe Sep verzerrter) verzerrter (bei hoher Sep)
	kombiniert	unbedeutend (bei geringer Sep: unverzerrter)	uneinheitlich (für hohe Sep verzerrter) verzerrter (bei hoher Sep)
Hohe Klassendistanz	weniger Verzerrung	weniger Verzerrung	weniger Verzerrung
Hohe Fallzahl	weniger Verzerrung	unbedeutend	unbedeutend
Kleine Klasse	unbedeutend	verzerrter (bei hoher Sep)	verzerrter (bei hoher Sep)
Varianzaufklärung der Simulationsgrößen Ausstrahlung auf Nachbarklassen	16 bis 52 % nicht sinnvoll identifizierbar	43 bis 60 % durch Mixing3 für Klasse 1	40 bis 60 % durch Mixing3 für Klasse 1

12 Zusammenfassung der Ergebnisse

Dieser Arbeit liegen drei allgemeine Forschungsfragen zu Grunde. Anhand dieser drei Forschungsfragen sollen im Weiteren auch die zentralen Ergebnisse der durchgeführten statistischen Simulation im Überblick zusammengefasst werden. Eine elaboriertere Ergebnisdarstellung kann den Kapiteln 9, 10 und 11 zu den einzelnen untersuchten Performanzkriterien sowie den dazugehörigen Überblickstabellen 10, 15, 26 und 40 auf Seite 112, 121, 147 und 169 entnommen werden.

- ad 1) Beeinflussen gruppenspezifische und längsschnittliche Verletzungen der Messinvarianz die Fähigkeit von MIGMMs die korrekte Anzahl latenter Wachstumsverlauffklassen zu identifizieren? Wenn ja, welche Faktoren beeinflussen die Stärke dieses Effekts?

In Abschnitt 6.2 wurde anhand der bisherigen Forschung die Arbeitshypothese aufgestellt, dass fehlende längsschnittliche Messinvarianz keinen starken verzerrenden Effekt auf die Klassenidentifikation haben sollte oder allenfalls bei geringer Klassendistanz ein Problem werden könnte. Für Messunterschiede zwischen den Gruppen wurde hingegen angenommen, dass die Klassenidentifikation beeinflusst wird. Dabei wurde auch darauf aufmerksam gemacht, dass fehlende Messinvarianz eine eigene Quelle von Heterogenität darstellt, welche die korrekte Klassenidentifikation ggf. sogar begünstigen kann. Des Weiteren wurde anhand erster Plausibilitätsüberlegungen angenommen, dass fehlerhafte Klassenidentifikationsraten durch fehlende Messinvarianz zusätzlich durch kleine Fallzahlen, geringe Klassengrößen und geringe Klassendistanzen verstärkt werden.

Wie aus der Ergebnisdarstellung in Kapitel 9.4 (und Kapitel 9.5.3) hervorgeht, muss tatsächlich bezüglich der Fähigkeit zur Klassenidentifikation deutlich zwischen längsschnittlicher und gruppenspezifischer Verletzung der Messinvarianz unterschieden werden. So konnte gezeigt werden, dass längsschnittliche Invarianzverletzungen tatsächlich, für sich genommen, die korrekte Klassenidentifikation nicht nennenswert beeinflussen. Dies gilt weitestgehend unabhängig vom zugrundeliegenden IC-Index.⁷⁹ Lediglich die Klassenidentifikation nach dem BLRT zeigte sich (bei hohen Klassendistanzen) geringfügig sensitiv gegenüber längsschnittlicher Non-Invarianz. Damit können die Ergebnisse von

⁷⁹Zwar konnten die vermuteten zusätzlichen Probleme bei geringen Fallzahlen nicht aufgedeckt werden, aber zumindest der AIC zeigt bei hoher Fallzahl für zeitliche Non-Invarianz eine leicht bessere Klassifikationsrate.

Zhang (2016) zur zeitlichen Messungsungleichheit bzgl. der korrekten Fallzuordnung und den Anteilswerten auch auf die Zielgröße Klassenidentifikation ausgeweitet werden.

Gruppenspezifische Messunterschiede wirken sich hingegen negativ auf die korrekte Klassenidentifikation aus. Während der AIC konstante (aber absolut schlechte) Identifikationsraten aufweist, reagieren der BIC und (in geringerem Umfang) der aBIC und BLRT gegenüber Gruppenunterschieden sensibel.

Moderierende Einflüsse von Fallzahl, Klassendistanz und Klassengrößen konnten anhand der Simulationsergebnisse nur teilweise bestätigt werden und widersprechen den ursprünglichen Erwartungen tlw. Während große Klassendistanz bei Gruppenunterschieden zu einer zusätzlichen Überidentifikation tendieren (da sich zwei Formen von Gruppenheterogenität aufaddieren), konnte für die Fallzahl keine einheitliche Moderationswirkung identifiziert werden. Kleine Klassengrößen führen bei Gruppenunterschieden sogar zu besseren Identifikationsraten, da die zusätzliche Gruppenheterogenität der Messungleichheit durch die kleine Gruppe teilweise kompensiert werden kann.

ad 2) Beeinflusst die Kombination aus gruppenspezifischer und längsschnittlicher Verletzung der Messinvarianz die Fähigkeit zur Klassenidentifikation zusätzlich? Wenn ja, welche Faktoren beeinflussen die Stärke dieses Effekts?

Aufgrund fehlender vorangegangener Untersuchungen hatten die Annahmen zur kombinierten Verletzung der Messinvarianz durch gruppen- und zeitspezifische Unterschiede einen starken explorativen Charakter. Die forschungsleitende Annahme war, dass gleichförmige Invarianzverletzungen (hinsichtlich der Vorzeichen der manifesten Intercepts und Faktorladungen) zu einer Verstärkung und entgegengesetzte Invarianzverletzungen ggf. zu einer Kompensation von Verzerrungen der Modellschätzung führen würden.

Hinsichtlich der Klassenidentifikation konnte gezeigt werden, dass die kombinierte Non-Invarianz im Ergebnis weder einheitlich zu einer wesentlich besseren oder schlechteren Identifikationsrate führt. Insbesondere für BIC, BLRT und den aBIC fällt die Verzerrung der Klassenidentifikation für die kombinierten Invarianzstufen insgesamt aber eher deutlicher aus. Aus den deskriptiven Auswertungen geht hervor, dass die Tendenz zur Überidentifikation von latenten Klassen aus den Invarianzstufen mit Gruppenunterschieden durch die Kombination mit den längsschnittlichen Invarianz gedreht wird und in eine Unteridentifikation latenter Klassen umschlägt. Dies führt aber im Ergebnis (wie aus den modellbasierten Analysen deutlich wird) nicht zwingend zu mehr korrekt identifizierten

Lösungen. Für moderierende Einflussgrößen gilt, dass zwar die schlechtere Klassenidentifikation von BIC und aBIC für die kombinierte Messinvarianzverletzung für kleine Klassengrößen und hohe Fallzahlen etwas abgemildert wird, aber hierdurch der Haupteffekt in der Summe nicht aufgehoben werden kann.

Das Simulationsergebnis macht deutlich, dass längsschnittliche Invarianzformen – für sich alleine genommen – keinen verzerrenden Einfluss auf die Klassenidentifikation ausüben⁸⁰, aber in Kombination mit Gruppenunterschieden durchaus relevant werden können und die Klassenidentifikation beeinflussen.

ad 3) In welchem Umfang lassen sich Parameterverzerrungen (Punktschätzer und Standardfehler) in den Wachstumsverlaufsfaktoren bei gruppenspezifischer, längsschnittlicher und kombinierter Verletzung der Messinvarianzannahme identifizieren? Welche Faktoren beeinflussen den Grad der Verzerrung zusätzlich?

Nach den Erwartungen aus Abschnitt 6.2 wurde davon ausgegangen, dass sowohl die längsschnittliche Verletzung der Messinvarianz als auch gruppenspezifische Abweichungen zu einer Verzerrung der Schätzung der Varianzen und der Parameterschätzer der durchschnittlichen Verlaufsformen führen. Auf Grundlage bisheriger Ergebnisse wurde darüber hinaus angenommen, dass für fehlende gruppenspezifische Messinvarianzen mit einer Typ-1-Fehlerinflation der Wachstumsparameter zu rechnen ist. Wenn auch eher spekulativ, wurde für die kombinierten Invarianzformen davon ausgegangen, dass bei gleichförmigen Invarianzverletzungen mit einer Verstärkung der Verzerrung zu rechnen ist und bei hinsichtlich der Stoßrichtung (bzw. Vorzeichen) entgegengesetzten Invarianzformen eine Kompensation zwischen längsschnittlicher und gruppenspezifischer Verzerrung angenommen werden kann. Für die weiteren Simulationsgrößen wie Klassendistanz, kleine Klassengröße und Fallzahl wurde, in Übereinstimmung mit den Ergebnissen aus der bisherigen GMM-Forschung, davon ausgegangen, dass große Distanzen, das Fehlen von besonders kleinen Klassen und eine hohe Fallzahl mit unverzerrteren Schätzungen verbunden sein sollten (vgl. Abschnitt 6.3).

Zunächst kann festgehalten werden, dass auch nach einer korrekten Klassenidentifikation die untersuchten Simulationsgrößen einen Einfluss auf die Schätzung der Wachstumsverlaufsform und die dazugehörigen Standardfehler haben.

⁸⁰Wenn man von den leichten Verzerrungen der Performanz des BLRT absieht.

Hinsichtlich des Bias der Wachstumsparameter hat sich gezeigt, dass die Fallzahl lediglich für die Varianzschätzungen relevant ist und die Mittelwertschätzungen der Wachstumsfaktoren dadurch weder negativ noch positiv beeinflusst werden. Für hohe Fallzahlen waren die Schätzungen der Varianzparameter der Wachstumsfaktoren weniger verzerrt. Das gleiche Muster trifft auch für die Verzerrung der Standardfehler zu. Die Fallzahl war für die Mittelwerte von Intercept- und Slopefaktor unbedeutend, wohingegen hohe Fallzahlen für die Varianzparameter zu weniger verzerrten SE-Schätzungen geführt haben.

Noch deutlicher fällt das Fazit für die Simulationsgröße Klassendistanz aus. Für alle untersuchten Wachstumsverlaufparameter fällt die Schätzung deutlich weniger verzerrt aus, wenn die Klassendistanz hoch ist. Identisch sieht es bei den Standardfehlerschätzungen aus. Sowohl die Standardfehler der Mittelwerte der Intercept- und Slopefaktoren als auch die Standardfehler der Varianzen wurden bei hoher Klassendistanz deutlich weniger verzerrt geschätzt.

Auch bei korrekter Klassenidentifikation wirken die verschiedenen Formen der Invarianzverletzungen auf die Performanz der Schätzung von MIGMMs. Für die zeitlichen non-invarianten Konfigurationen lässt sich zeigen, dass sie nur dann einen bedeutenden verzerrenden Effekt auf den Bias der Varianzen-, Intercept- und Slopeparameter haben, wenn die Faktorladungen nicht-invariant generiert wurden. Bzgl. der Standardfehlerschätzungen zeigt sich, dass die Kovarianzen für die zeitlichen Konfigurationen nicht betroffen sind. Die Standardfehler für Intercept und Slopes werden hingegen, insbesondere bei großer Klassendistanz, auch bei lediglich zeitlicher Invarianzverletzung, verzerrter geschätzt.

Messinvarianzverletzungen zwischen den latenten Klassen wirken hingegen einheitlich verzerrend auf den Bias der Varianzen, Intercepts und Slopes. Für die Varianzparameter und die Slopes gilt dies insbesondere für hohe Klassendistanzen. Die Standardfehlerschätzungen sind für gruppenspezifische Invarianzverletzungen für die Varianzen entweder unbedeutend oder sogar weniger verzerrt. Ein anderes Bild zeigt sich für die Standardfehler der Intercepts und Slopes. Für hohe Klassendistanzen werden diese im Schnitt verzerrter geschätzt.

Die Kombination aus längsschnittlicher und gruppenspezifischer Invarianzverletzung zeigt hinsichtlich des Bias ein gemischtes Bild. Für die Varianzschätzungen zeigen sich starke/verstärkte Verzerrungen, besonders bei hohen Klassendistanzen. Die Slopewerte sind hingegen bei hohen Klassenunterschieden weniger verzerrt als die reinen gruppenspezifischen Konfigurationen, was auf einen gewissen Kompensationseffekt durch die beiden

Invarianzformen hinweist, ohne dabei jedoch gänzlich unverzerrt zu sein. Die Biaswerte für die Intercepts sind hingegen vergleichbar verzerrt zu den gruppenspezifischen Konfigurationen. Die Verzerrungen der Standardfehlerschätzungen weisen hingegen das gleiche Muster auf, wie für die gruppenspezifischen Invarianzen: keine relevanten Verzerrungen für die Varianzparameter und eine verzerrtere Schätzung für Intercepts und Slopes bei hoher Klassendistanz.

Insgesamt muss festgehalten werden, dass die Performanz von MIGMMs auch über die Klassenidentifikation hinaus durch die Simulationsgrößen beeinflusst wird und somit auch die Rekonstruktion der Wachstumsverlaufsform unter gewissen Bedingungen gefährdet ist, auch wenn die latente Klassenanzahl zuvor korrekt bestimmt wurde. Besonders ist hierbei die Klassendistanz hervorzuheben. Der mit dieser Größe verbundene Haupteffekt hat über alle Zielgrößen hinweg einen starken und konsistenten Einfluss auf die Modellperformanz. Zusätzlich moderiert er auch die Bedeutsamkeit weiterer Simulationsgrößen wie die der Invarianzformen auf die Modellperformanz hinsichtlich des Bias der Parameterschätzungen und der Standardfehler.

13 Diskussion der Leistungsfähigkeit von MIGMMs für sozialwissenschaftliche Analysen

Im Kern geht diese Arbeit der Frage nach, ob MIGMMs gewinnbringend für die empirische Sozialforschung eingesetzt werden können oder unter welchen Bedingungen der Einsatz von MIGMMs zu vertreten ist. Die Ergebnisse dieser Studie machen klar, dass ein naiver und unreflektierter Einsatz der Analysemethodik zu stark verzerrten Ergebnissen führen kann. Diese Verzerrungen betreffen sowohl das genuine Einsatzziel der latenten Klassenidentifikation als auch die daran anschließende Rekonstruktion der Wachstumsverlaufsmuster anhand der Kennzahlen der latenten Wachstumsparameter. Es konnte somit gezeigt werden, dass ungünstige Datensituationen sowohl die Identifikation der den Daten zugrundeliegenden Klassen anhand der verwendeten Indizes gefährden können als auch die durchschnittlichen Verlaufsformen der Entwicklung der betrachteten Zielvariable sowie den Referenzschluss auf die Grundgesamtheit verzerren. Auch wenn diese Schlussfolgerungen einen gewissen Pessimismus bzgl. der empirischen Anwendungen vermitteln, lassen sich konstruktive Aspekte herausarbeiten, welche zwar keinen blinden Optimismus für die Nutzung des Verfahrens befördern, aber dennoch Bedingungen aufzeigen, unter

welchen die Modellschätzungen verlässlich und sinnvoll eingesetzt werden können. Vor dem Hintergrund der präsentierten Ergebnisse werden folgende Orientierungshilfen (bzw. Warnungen) für den Einsatz von MIGM-Modellierungen abgeleitet:

- **Klassendistanz:** Nur bei deutlichen Unterschieden bzgl. der Verlaufsmuster scheinen MIGMMs zuverlässige Ergebnisse zu liefern. Werden schon aus einer analytisch-theoretischen Perspektive geringe Unterschiede erwartet, kann den Kennzahlen der Klassenidentifikation und den geschätzten Verlaufsmustern mit keinem ausreichenden Sicherheitsniveau vertraut werden. Die explorative Qualität/Zuverlässigkeit von MIGMMs stößt bei geringen Klassenunterschieden an ihre Grenzen.
- **Messkonstrukte:** Messunterschiede zwischen den latenten Klassen führen zu den deutlichsten Verzerrungen hinsichtlich der Klassenidentifikation und der Parameterschätzungen. Daher sollten die analysierten Messkonstrukte für MIGMMs möglichst zuverlässig hinsichtlich Messäquivalenz sein. Dies lässt sich nur plausibel für Messkonstrukte unterstellen, welche zumindest bereits anhand bekannten Gruppen Messinvarianz gezeigt haben, um die Messinvarianz auch hinsichtlich unbekannter Gruppen wahrscheinlicher zu machen. Ein Verzicht auf MIGMMs durch eine Schätzung von GMMs ohne latente Messung (nur mit Composite Score Werten) stellt keine Lösung des Problems dar, da die verzerrenden Effekte implizit mitgetragen werden. Fehlende Messinvarianz über die Zeit hinweg scheint zwar weniger starke Auswirkungen zu haben, zeigt aber in Kombination mit Gruppenunterschieden deutliche Verzerrungen. Daher kann eine vorangehende Überprüfung – vor der Schätzung des eigentlichen MIGMMs – dennoch angeraten werden.
- **Klassenidentifikation:** Leider existiert keine einzelne Maßzahl, welche die beste Identifikationsleistung bzgl. der latenten Klassen unter allen Daten- und Messsituationen gewährleisten kann. Der BIC schneidet für wünschenswerte Datensituationen wie große Klassendistanz und vorhandene Messinvarianz sehr gut ab. Der AIC kann zumindest für unklare Daten- und Messsituationen eine Orientierung für die maximale Anzahl an latenten Klassen liefern. Der aBIC und (in Teilen) der BLRT sind die Maß der Wahl, wenn von ungünstigen Datensituationen ausgegangen werden kann oder der/die Forscher/Forscherin sich nicht ausreichend sicher über die Daten- und Messsituation ist. Dennoch stellt keines der untersuchten Kriterien einen hin-

reichenden potentiellen Falsifikator (oder Verifikation) für begründete analytische oder theoretische Annahmen dar.

Einschränkend muss an dieser Stelle noch einmal darauf hingewiesen werden, dass die Ergebnisse dieser Studie – und die daraus gezogenen Schlussfolgerungen – nicht zwangsläufig auf Datensituation übertragen werden können, welche sich nicht im Rahmen der hier untersuchten Simulationskonfigurationen bewegen. Dabei handelt es sich um eine generelle, der Methodik der Monte-Carlo-Simulation inhärente, Beschränkung, welche prinzipiell sowohl nicht manipulierte Modellgrößen, nicht explizit modellierte Kategorienausprägungen der Simulationsvariablen und Wertebereiche außerhalb des untersuchten Wertebereichs betrifft. Um Komplexität zu reduzieren, Nachvollziehbarkeit zu vereinfachen und handhabbare Simulationslaufzeiten zu gewährleisten, wurde in der durchgeführten Simulation z. B. darauf verzichtet komplexere Wachstumsverlaufsmuster zu modellieren und lediglich ein linearer Wachstumsverlauf angenommen. Auch die manifesten Variablen wurden als metrisch spezifiziert und die Klassendistanz über eine Kombination aus Unterschieden in den Intercept- und Slopemittelwerten zwischen den Klassen hergestellt. Vor dem Hintergrund der Ergebnisse der Simulation scheinen vor allem zwei Aspekte bzgl. dieser Beschränkungen herauszustechen. Damit sind die Klassendistanz und (zu einem geringeren Umfang) die Fallzahl gemeint. Die damit verbundenen offenen Fragen und Anknüpfungspunkte für weitere Analysen zur Performanz von MIGMMs sollen u. a. im Folgenden diskutiert werden.

13.1 Desiderate für nachfolgende Analysen

Eines der konsistenten Ergebnisse dieser Untersuchung ist die Bedeutung der Klassendistanz für alle betrachteten Performanzkriterien von MIGMMs. Besonders frappierend war dabei i. d. R. der Unterschied zwischen den Ausprägungen 1.5 MD und 4.0 MD. Während die größere Klassendistanz mit deutlich besseren Klassenidentifikationsraten und deutlich weniger verzerrten Parameterschätzungen und Standardfehlern assoziiert war, zeigte sich dies bereits in der Ausprägung von 1.5 MD nicht mehr. Obwohl gängige Einschätzungen aus der GMM-Literatur bei einer Distanz von 1.5 MD von «großen» Unterschieden ausgehen, erscheint dies für den MIGMM-Kontext, zumindest nach den hier präsentierten Ergebnissen, nicht mehr als haltbar. Weitere Untersuchungen sollten daher versuchen zu bestimmen, wo genau zwischen der Spannbreite von 1.5 und 4.0 MD eine akzeptable

Modellschätzung für MIGMMs möglich ist. Damit verbunden ist auch die Frage danach, ob die Performanz von MIGMMs nur eine Funktion der absoluten multivariaten Distanz darstellt oder ob die Art der Herstellung der Distanz (z. B. nur über Intercepts, über Varianzen der Wachstumsparameter oder eine Kombination aus Intercepts und Slopes) ebenfalls eine Rolle spielt. Dies ist nicht zuletzt deswegen ein Forschungsdesiderat, da innerhalb der Simulationsliteratur keine Einigkeit über die passendste Herstellung von Klassendifferenz besteht (vgl. hierzu Abschnitt 3.2.2).

Der BLRT hat sich hinsichtlich der Klassenidentifikation für MIGMMs nicht als substantiell besseres (oder schlechteres) Maß als die IC-Indizes erwiesen. Wie bereits an entsprechender Stelle erwähnt (vgl. Abschnitt 9.5.2), kann dieses Ergebnis evtl. auch darauf zurückgeführt werden, dass die Anzahl der Bootstrap-Ziehungen per Mplus Defaultoption nach einem sequenziellen Verfahren erfolgte und damit die Performanz nicht optimal ausfallen konnte. Die Entscheidung für die weniger rechenintensive Defaultoption ist dabei nicht aufgrund von analytischen Überlegungen gefallen, sondern war rein forschungspraktisch (aufgrund der Rechenzeitressourcen) motiviert. Allerdings bemerken auch McNeish & Harring (2017), dass der BLRT in Simulationsstudien oft gänzlich übergangen wird, da er mit sehr großem zusätzlichem Rechenaufwand verbunden ist (vgl. McNeish & Harring, 2017: 239). Für den Anwendungsbereich der MIGMMs wären weitere Performanzevaluationen wünschenswert, welche die Ziehungszahl auf eine möglichst hohe Zahl (z. B. 1000) fixieren könnten. Damit könnte die in dieser Studie gefundene Leistungsfähigkeit weiter abgesichert werden. Für eine systematische Untersuchung im Rahmen einer Monte-Carlo-Simulation wäre hierfür allerdings eine deutlich leistungstärkere Rechnerinfrastruktur nötig, welche nicht nur auf Parallelisierung der Berechnung, sondern auch auf Rechenzeit bzw. Rechengeschwindigkeit ausgelegt ist.

Die Simulationsgröße «Fallzahl» erwies sich in den Analysen nicht als besonders einflussreiche Größe für die untersuchten Performanzkriterien. Lediglich im Zusammenhang mit besonders kleinen Klassenanteilen und für die Varianzparameterschätzungen konnten relevante Einflüsse gefunden werden. Dies mag daran liegen, dass die kleinste Kategorienausprägung mit 300 Fällen noch zu groß gewählt wurde, um eine kritische Schwelle zu übertreten.⁸¹ Zumindest einzelne Auswertungen lassen vermuten, dass die Performanz der

⁸¹Die Festlegung der Kategorien war in dieser Studie besonders durch die zusätzliche Simulationsgröße «Klassenanteile» (bzw. kleine Klassen) motiviert, da in den Simulationsdaten bei Klassenanteilen von 5 % (in Kombination mit der Fallzahl von 300) immerhin noch durchschnittlich 15 Fälle der kleinsten Klasse angehören.

Modelle bei kleineren Fallzahlen eingebrochen wäre, ohne dass jedoch bei 300 Fällen bereits die relevanten Grenzwerte überschritten wurden. Daher sollte nicht der Fehlschluss entstehen, dass die Fallzahl keine besonders relevante Größe für die Modellierung von MIGMMs darstellt. Vielmehr gilt es in weiteren Studien und Analysen kritische Schwellenwerte auch für den Kontext von MIGMMs zu bestimmen um Empfehlungen für die empirische Praxis zu präzisieren.

Eine alternative Möglichkeit des Umgangs mit fehlender Messinvarianz innerhalb von MIGMMs stellt die prinzipiell mögliche freie Schätzung der Messstruktur innerhalb der latenten Klassen dar. Kim & Wang (2017) konnten für latente Gruppenmessdifferenzen zeigen, dass die Wachstumsverläufe bei derlei Schätzungen unverzerrter sind als invariante Spezifikationen. Allerdings ergeben sich dabei sowohl praktische als auch konzeptuelle Schwierigkeiten. Aus der praktischen Perspektive werden die Schätzmodelle hierbei mit zusätzlicher Komplexität in Form von freien Parametern konfrontiert, was die Konvergenz der Modelle, insbesondere für empirisch generierte Daten, weiter erschwert. Bereits invariante Spezifikationen müssen sich mit einfacheren Modellierungen begnügen, ohne dass zusätzlich die Messstruktur der latenten Klassen einzeln bestimmt werden muss (vgl. z. B. Krause et al., 2019: 400 ff.). Konzeptionell stellt sich bei derlei Modellierungen die Frage, ob ein Vergleich der Verlaufsmuster von latenten Größen und insbesondere die Extraktion von latenten Klassen anhand dieser Verlaufsmuster, welche weder eine äquivalente Verwendung der Skala noch eine gleiche Zusammensetzung des Konstrukts haben, noch ein sinnvolles Unterfangen darstellt. Inwieweit diese Probleme für die Praxis relevant sind und inwieweit ein Trade-off zwischen Vergleichbarkeit und Unverzerrtheit möglich ist, muss allerdings in weiteren Arbeiten umfassender evaluiert werden.

Da der Schwerpunkt dieser Arbeit auf der Messinvarianz innerhalb von MIGMMs sowie auf der Übertragung von Erkenntnissen bereits etablierter Einflussgrößen lag, wurde bei der Datengenerierung der manifesten Variablen von einem einfachen metrischen Messniveau ausgegangen. Für Arbeiten mit einem anderen inhaltlichem Fokus könnte sich auch die explizite Modellierung von quasi-metrischen Messskalen (bzw. ordinalen Antwortskalen) lohnen, welche lediglich zwischen 4 bis 10 Antwortkategorien haben und somit Daten aus der empirischen Praxis der quantitativen Sozialforschung näher kommen als eine rein metrische Datengenerierung wie in der vorliegenden Arbeit. Allgemein handelt es sich dabei um eine bisher noch nicht ausreichend beachtete Datensituation für Growth Mixture Modelle, was für den Spezialfall der MIGMMs noch einmal deutlicher ausfällt.

13.2 Resümee

Ob Multiple Indicator Growth Mixture Modelle das sozialwissenschaftliche Analyserepertoire für längsschnittliche Untersuchungen tatsächlich sinnvoll ergänzen können, kann nur die Zukunft ihrer praktischen empirischen Anwendung zeigen. Erste Beiträge wie dieser, auf Grundlage von mathematisch-statistischen Analysen, könnten zum einen zeigen, unter welchen Bedingungen der Anspruch der datengeleiteten Identifikation von latenten Gruppenverläufen bei gleichzeitiger Berücksichtigung von latenten Messkonstrukten tatsächlich eingelöst werden kann. Vielleicht noch viel wichtiger für die Anwendungspraxis scheint aber die explizite (und durch quantitative Maße gestützte) Darlegung der Bedingungen, unter welchen der Sozialforscher oder die Sozialforscherin sich nicht auf die daten- und kennzahlgeleitete Struktur- und Musterentdeckung verlassen kann und sich auf den explorativen Charakter von Mixture-Modellen zurückbesinnen muss um Fehler erster und zweiter Art zu vermeiden. Nur bei Berücksichtigung der Grenzen und Fallstricken von immer elaborierten statistischen Analyseverfahren kann letztlich sichergestellt werden, dass ein inhaltlicher Mehrwert für die Forschung genutzt werden kann, ohne dass mess- oder analysetechnische Artefakte diese Vorteile schmälern oder gänzlich zunichtemachen.

14 Literatur

- Aitkin, M. & Rubin, D. B. (1985). Estimation and hypothesis testing in finite mixture models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(1):67–75.
- Bandalos, d. L. & Gagne, P. (2012). Simulation methods in structural equation modeling. In Hoyle, R. H., editor, *Handbook of structural equation modeling*, pages 92–108. Guilford Press, New York.
- Bandalos, d. L. & Leite, W. (2013). Use of monte carlo studies in structural equation modeling research. In Hancock & Gregory R., editors, *Structural equation modeling, Quantitative methods in education and the behavioral sciences : issues, research, and teaching*, pages 625–666. Information Age Publishing, Inc., Charlotte, NC.
- Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42(4):757–786.
- Bauer, D. J. & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8(3):338–363.
- Bauer, D. J. & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, 9(1):3–29.
- Berlin, K. S., Parra, G. R., & Williams, N. A. (2014). An introduction to latent variable mixture modeling (part 2): longitudinal latent class growth analysis and growth mixture models. *Journal of Pediatric Psychology*, 39(2):188–203.
- Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):719–725.
- Biernacki, C. & Govaert, G. (1997). Using the classification likelihood to choose the number of clusters. *Computing Science and Statistics*, pages 451–457.
- Boomsma, A., Hoyle, R. H., & Panter, A. T. (2012). The structural equation modeling research report. In Hoyle, R. H., editor, *Handbook of structural equation modeling*, pages 341–358. Guilford Press, New York.

- Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis: Chapter 7: Converting Among Effect Sizes*. Wiley, Chichester, West Sussex, UK.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44(11):S. 176–181.
- Boscardin, C. K., Muthén, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology*, 100(1):192–208.
- Bozdogan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Methodology in the social sciences. Guilford Press, New York.
- Callina, K. S., Johnson, S. K., Buckingham, M. H., & Lerner, R. M. (2014). Hope in context: developmental profiles of trust, hopeful future expectations, and civic engagement across adolescence. *Journal of Youth and Adolescence*, 43(6):869–883.
- Carsey, T. M. & Harden, J. J. (2014). *Monte Carlo Simulation and Resampling Methods for Social Science*. Sage Publications, Los Angeles, Calif.
- Celeux, G. & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2):195–212.
- Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (lmcs) and multiple indicator latent growth modeling (mlgm). *Organizational Research Methods*, 1(4):421–483.
- Chen, Q., Kwok, O.-M., Luo, W., & Willson, V. L. (2010). The impact of ignoring a level of nesting structure in multilevel growth mixture models: A monte carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(4):570–589.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates, Hillsdale, N.J., 2nd ed. edition.

- Davidson, R. & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. Oxford University Press, New York.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.
- Depaoli, S. (2013). Mixture class recovery in gmm under varying degrees of class separation: frequentist versus bayesian estimation. *Psychological Methods*, 18(2):186–219.
- Diallo, T. M. O., Morin, A. J. S., & Lu, H. (2016). Impact of misspecifications of the latent variance–covariance and residual matrices on the class enumeration accuracy of growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(4):507–531.
- Diallo, T. M. O., Morin, A. J. S., & Lu, H. (2017). The impact of total and partial inclusion or exclusion of active and inactive time invariant covariates in growth mixture models. *Psychological Methods*, 22(1):166–190.
- Duncan, T. E. & Duncan, S. C. (2009). The abc’s of lgm: An introductory guide to latent variable growth curve modeling. *Social and Personality Psychology Compass*, 3(6):979–991.
- Duncan, T. E., Duncan, S. C., Strycker, L. A., Okut, H., & Li, F. (2002). Growth mixture modeling of adolescent alcohol use data. Retrieved from Oregon Research Institute Web site: <http://www.ori.org/methodology>.
- Enders, C. K. & Tofghi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(1):75–95.
- Falk, M., Hain, J., Marohn, F., Fischer, H., & Michel, R., editors (2014). *Statistik in Theorie und Praxis*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C.-P., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshmen college students. *Journal of Consulting and Clinical Psychology*, 73(2):229–238.

- Grimm, K. J., Mazza, G. L., & Davoudzadeh, P. (2017). Model selection in finite mixture models: A k -fold cross-validation approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(2):246–256.
- Grimm, K. J. & Ram, N. (2009). A second-order growth mixture model for developmental research. *Research in Human Development*, 6(2-3):121–143.
- Hancock, G. R. & Samuelsen, K. M., editors (2008). *Advances in latent variable mixture models*. The CILVR series on latent variable methodology. Information Age Pub, Charlotte, NC.
- Hayes, A. F. & Cai, L. (2007). Using heteroskedasticity-consistent standard error estimators in ols regression: An introduction and software implementation. *Behavior Research Methods*, 39(4):709–722.
- He, J. & Fan, X. (2018). Evaluating the performance of the k -fold cross-validation approach for model selection in growth mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(3):1–14.
- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(2):202–226.
- Hoogland, J. J. & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods & Research*, 26(3):329–367.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Quantitative methodology series. Routledge, New York, 2nd ed. edition.
- Hu, J., Leite, W. L., & Gao, M. (2017). An evaluation of the use of covariates to assist in class enumeration in linear growth mixture modeling. *Behavior Research Methods*.
- James, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, 41(1/2):19–43.
- Kim, E. S. & Wang, Y. (2017). Class enumeration and parameter recovery of growth mixture modeling and second-order growth mixture modeling in the presence of measurement noninvariance between latent classes. *Frontiers in Psychology*, 8:1499.

- Kim, E. S. & Willson, V. L. (2014). Measurement invariance across groups in latent growth modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3):408–424.
- Kim, S.-Y. & Kim, J.-S. (2012). Investigating stage-sequential growth mixture models with multiphase longitudinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(2):293–319.
- Kirves, K., Kinnunen, U., de Cuyper, N., & Mäkikangas, A. (2014). Trajectories of perceived employability and their associations with well-being at work. *Journal of Personnel Psychology*, 13(1):46–57.
- Koo, N. & Leite, W. L. (2014). The impact of ignoring time series processes in linear growth mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2):210–224.
- Krause, T., Wahl, A., & Wuketich, M. (2019). Wachstumsverläufe von latenten klassen in der einstellungs- und verhaltensforschung: Am beispiel der sozialen exklusion von kindern und jugendlichen in deutschland. In Mayerl, J., Krause, T., Wahl, A., & Wuketich, M., editors, *Einstellungen und Verhalten in der empirischen Sozialforschung*, pages 387–413. Springer VS, Wiesbaden.
- Leite, W. L. (2007). A comparison of latent growth models for constructs measured by multiple items. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4):581–610.
- Leuchter, A. F., Cook, I. A., Witte, E. A., Morgan, M., & Abrams, M. (2002). Changes in brain function of depressed subjects during treatment with placebo. *The American Journal of Psychiatry*, 159(1):122–129.
- Li, F., Barrera, M., Hops, H., & Fisher, K. J. (2002). The longitudinal influence of peers on the development of alcohol use in late adolescence: a growth mixture analysis. *Journal of Behavioral Medicine*, 25(3):293–315.
- Li, L. & Hser, Y.-I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research*, 46(2):266–302.
- Li, M., Harring, J. R., & Macready, G. B. (2014). Investigating the feasibility of using mplus in the estimation of growth mixture models. *Journal of Modern Applied Statistical Methods*, 13(1):484–513.

- Liu, M. & Hancock, G. R. (2014). Unrestricted mixture models for class identification in growth mixture modeling. *Educational and Psychological Measurement*, 74(4):557–584.
- Lo, Y. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3):767–778.
- Lubke, G. & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research*, 41(4):499–532.
- Magidson, J. & Vermunt, J. K. (2004). Latent class models. *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, pages 175–198.
- Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In *The Oxford handbook of quantitative methods: Statistical analysis, Vol. 2*, Oxford library of psychology, pages 551–611. Oxford University Press, New York, NY, US.
- McArdle, J. J. & Epstein, D. B. (1987). Latent growth curves within developmental structural equation models. *Child Development*, 58(1):110–133.
- McArdle, J. J. & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. *Handbook of Psychology*.
- McLachlan, G. J. & Peel, D. (2000). *Finite mixture models*. Wiley series in probability and statistics. Applied probability and statistics section. Wiley, New York.
- McNeish, D. & Harring, J. R. (2017). The effect of model misspecification on growth mixture model class enumeration. *Journal of Classification*, 52(2):226–248.
- Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1):107–122.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge, New York.
- Mooney, C. Z. (1997). *Monte Carlo simulation*, volume 116. Sage Publications, Thousand Oaks, CA [u.a.].
- Morin, A. J. S., Bujacz, A., & Gagné, M. (2018). Person-centered methodologies in the organizational sciences. *Organizational Research Methods*, page 109442811877385.

- Morin, A. J. S., Maïano, C., Nagengast, B., Marsh, H. W., Morizot, J., & Janosz, M. (2011). General growth mixture analysis of adolescents' developmental trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(4):613–648.
- Muthén, B. (2001). Latent variable mixture modeling. *New Developments and Techniques in Structural Equation Modeling*, pages 1–33.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In Kaplan, D., editor, *The SAGE handbook of quantitative methodology for the social sciences*, pages 345–368. SAGE, Thousand Oaks, Calif. and London.
- Muthén, B. & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-gaussian random effects. *Longitudinal data analysis*, pages 143–165.
- Muthén, B., Asparouhov, T., Hunter, A. M., & Leuchter, A. F. (2011). Growth modeling with nonignorable dropout: alternative analyses of the star*d antidepressant trial. *Psychological Methods*, 16(1):17–33.
- Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the em algorithm. *Biometrics*, 55(2):463–469.
- Muthén, L. K. & Muthén, B. (2015). Mplus user's guide: 7.3: https://www.statmodel.com/download/usersguide/mplus%20user%20guide%20ver_7_r3_web.pdf.
- Muthén, L. K. & Muthén, B. O. (2002). How to use a monte carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4):599–620.
- Muthén & Muthén (2003). Perturbation of starting values: Technical appendices: <https://www.statmodel.com/download/starts.pdf>.
- Nagin, D. & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development*, 70(5):1181–1196.

- Nagin, D. S. (2010). Group-based trajectory modeling: An overview. In Piquero, A. R. & Weisburd, D., editors, *Handbook of Quantitative Criminology*, pages 53–67. Springer New York, New York, NY.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4):535–569.
- Olivera-Aguilar, M. (2013). *Impact of Violations of Longitudinal Measurement Invariance in Latent Growth Models and Autoregressive Quasi-simplex Models*. PhD thesis, Arizona State University.
- O.V. (2017). bwunicluster: <http://www.scc.kit.edu/dienste/bwunicluster.php>.
- Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in english and finnish. *Journal of Educational Psychology*, 97(3):299–319.
- Petras, H. & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In Piquero, A. R. & Weisburd, D., editors, *Handbook of Quantitative Criminology*, pages 69–100. Springer Science+Business Media, LLC, New York, NY.
- Peugh, J. & Fan, X. (2012). How well does growth mixture modeling identify heterogeneous growth trajectories? a simulation study examining gmm’s performance characteristics. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(2):204–226.
- Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*, volume 157 of *Quantitative applications in the social sciences*. SAGE, Thousand Oaks.
- Qureshi, I. & Fang, Y. (2010). Socialization in open source software projects: A growth mixture modeling approach. *Organizational Research Methods*, 14(1):208–238.
- Ram, N. & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, 33(6):565–576.

- Reinecke, J. (2012). *Wachstumsmodelle*. Sozialwissenschaftliche Forschungsmethoden. Rainer Hampp Verlag, Mering.
- Reinecke, J. (2014). Wachstums- und mischverteilungsmodelle in der kriminologischen längsschnittforschung. In Eifler, S. & Pollich, D., editors, *Empirische Forschung über Kriminalität*, pages 439–474. Springer Fachmedien Wiesbaden, Wiesbaden.
- Reinecke, J., Meyer, M., & Boers, K. (2015). Stage-sequential growth mixture modeling of criminological panel data. In Stemmler, M., von Eye, A., & Wiedermann, W., editors, *Dependent data in social sciences research*, volume 145 of *Springer Proceedings in Mathematics & Statistics*, pages 67–89. Springer, Cham.
- Reinecke, J. & Seddig, D. (2011). Growth mixture models in longitudinal research. *AStA Advances in Statistical Analysis*, 95(4):415–434.
- Rogosa, D. R. & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, 50(2):203–228.
- Rupp, A. A. (2013). Clustering and classification. In *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52(3):333–343.
- Serang, S., Grimm, K. J., & McArdle, J. J. (2016). Estimation of time-unstructured nonlinear mixed-effects mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, pages 1–14.
- Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press, New York and Oxford and New York and Oxford.
- Stoolmiller, M., Kim, H. K., & Capaldi, D. M. (2005). The course of depressive symptoms in men from early adolescence to young adulthood: identifying latent trajectories and early predictors. *Journal of Abnormal Psychology*, 114(3):331–345.
- Tekle, F. B., Gudicha, D. W., & Vermunt, J. K. (2016). Power analysis for the bootstrap likelihood ratio test for the number of classes in latent class models. *Advances in Data Analysis and Classification*, 10(2):209–224.

- Tofighi, D. & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In Hancock, G. R. & Samuelsen, K. M., editors, *Advances in latent variable mixture models*, The CILVR series on latent variable methodology, pages 317–341. Information Age Pub, Charlotte, NC.
- Tolvanen, A. (2007). *Latent growth mixture modeling: A simulation study*, volume 111 of *Report / University of Jyväskylä, Department of Mathematics and Statistics Bericht / Universität Jyväskylä, Institut für Mathematik und Statistik*. University of Jyväskylä, Dept. of Mathematics and Statistics, Jyväskylä.
- Tueller, S. & Lubke, G. (2010). Evaluation of structural equation mixture models parameter estimates and correct class assignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 17(2):165–192.
- Tueller, S. J., Drotar, S., & Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(1):110–131.
- Urban, D. (2004). *Neue Methoden der Längsschnittanalyse: Zur Anwendung von latenten Wachstumskurvenmodellen in Einstellungs-und Sozialisationsforschung*. Lit Verlag, Münster.
- Urban, D. & Mayerl, J. (2014). *Strukturgleichungsmodellierung*. Springer Fachmedien Wiesbaden, Wiesbaden.
- Urban, D. & Mayerl, J. (2018). *Angewandte Regressionsanalyse: Theorie, Technik und Praxis*. Studienskripten zur Soziologie. Springer Fachmedien Wiesbaden, Wiesbaden.
- Usami, S., Hayes, T., & McArdle, J. (2017). Fitting structural equation model trees and latent growth curve mixture models in longitudinal designs: The influence of model misspecification. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4):585–598.
- Wang, S., Chen, C.-C., Dai, C.-L., & Richardson, G. B. (2018). A call for, and beginner’s guide to, measurement invariance testing in evolutionary psychology. *Evolutionary Psychological Science*, 4(2):166–178.
- Wickrama, K. A. S. (2016). *Higher-order growth curves and mixture modeling with Mplus: A practical guide*. Routledge, New York, NY.

- Windle, M. & Wiesner, M. (2004). Trajectories of marijuana use from adolescence to young adulthood: Predictors and outcomes. *Development and Psychopathology*, 16(04).
- Wirth, R. J. (2008). *The effects of measurement non-invariance on parameter estimation in latent growth models*. Dissertation, The University of North Carolina at Chapel Hill, Chapel Hill.
- Wu, J. & Witkiewitz, K. (2008). Network support for drinking: an application of multiple groups growth mixture modeling to examine client-treatment matching. *Journal of Studies on Alcohol and Drugs*, 69(1):21–29.
- Zhang, C. (2016). *Longitudinal Measurement Non-Invariance on Growth Parameters Recovery and Classification Accuracy in Growth Mixture Modeling*. Dissertation, University of Pittsburgh, Pittsburgh.

A Anhang

A.1 Tabellen und Abbildungen zur Klassenidentifikation

Tabelle 41: Prozentuale Klassenidentifikation nach allen Konfigurationsparametern (Referenz).

	invar	mixing	sep	nobs	1.aic	2.aic	3.aic	4.aic	1.bic	2.bic	3.bic	4.bic	1.abic	2.abic	3.abic	4.abic
1	1.00	1.00	1.00	1.00	0.67	79.33	17.00	3.00	22.67	77.33	0.00	0.00	1.33	89.67	8.00	1.00
2	1.00	1.00	1.00	2.00	0.00	81.67	13.33	5.00	0.33	99.67	0.00	0.00	0.00	98.00	1.67	0.33
3	1.00	1.00	1.00	3.00	0.00	75.67	20.67	3.67	0.00	100.00	0.00	0.00	0.00	99.00	1.00	0.00
4	1.00	1.00	2.00	1.00	0.00	75.33	20.33	4.33	0.00	100.00	0.00	0.00	0.00	88.00	10.33	1.67
5	1.00	1.00	2.00	2.00	0.00	67.33	24.33	8.33	0.00	100.00	0.00	0.00	0.00	95.33	4.67	0.00
6	1.00	1.00	2.00	3.00	0.00	60.33	30.00	9.67	0.00	100.00	0.00	0.00	0.00	96.00	4.00	0.00
7	1.00	1.00	3.00	1.00	0.00	0.00	61.33	38.67	0.00	1.00	98.33	0.67	0.00	0.00	73.33	26.67
8	1.00	1.00	3.00	2.00	0.00	0.00	60.00	40.00	0.00	0.00	99.67	0.33	0.00	0.00	89.33	10.67
9	1.00	1.00	3.00	3.00	0.00	0.00	64.33	35.67	0.00	0.00	100.00	0.00	0.00	0.00	95.00	5.00
10	1.00	2.00	1.00	1.00	5.33	72.67	19.33	2.67	56.67	43.33	0.00	0.00	6.67	84.33	8.00	1.00
11	1.00	2.00	1.00	2.00	0.00	76.92	18.39	4.68	7.36	92.64	0.00	0.00	0.67	95.99	2.68	0.67
12	1.00	2.00	1.00	3.00	0.00	74.33	22.00	3.67	0.33	99.67	0.00	0.00	0.00	99.33	0.67	0.00
13	1.00	2.00	2.00	1.00	0.00	79.00	13.33	7.67	0.67	99.33	0.00	0.00	0.00	91.67	5.67	2.67
14	1.00	2.00	2.00	2.00	0.00	75.67	19.00	5.33	0.00	100.00	0.00	0.00	0.00	95.00	4.67	0.33
15	1.00	2.00	2.00	3.00	0.00	70.33	19.00	10.67	0.00	100.00	0.00	0.00	0.00	98.00	2.00	0.00
16	1.00	2.00	3.00	1.00	0.00	0.00	61.00	39.00	0.00	9.00	89.33	1.67	0.00	0.00	74.67	25.33
17	1.00	2.00	3.00	2.00	0.00	0.00	62.00	38.00	0.00	0.00	99.67	0.33	0.00	0.00	89.33	10.67
18	1.00	2.00	3.00	3.00	0.00	0.00	62.67	37.33	0.00	0.00	100.00	0.00	0.00	0.00	95.67	4.33
19	1.00	3.00	1.00	1.00	2.00	76.33	19.67	2.00	40.67	59.33	0.00	0.00	3.67	86.33	9.00	1.00
20	1.00	3.00	1.00	2.00	0.00	76.00	18.00	6.00	3.67	96.33	0.00	0.00	0.00	96.33	3.33	0.33
21	1.00	3.00	1.00	3.00	0.00	76.67	18.33	5.00	0.00	100.00	0.00	0.00	0.00	99.33	0.67	0.00
22	1.00	3.00	2.00	1.00	0.00	81.67	14.33	4.00	0.00	100.00	0.00	0.00	0.00	92.67	6.00	1.33
23	1.00	3.00	2.00	2.00	0.00	86.33	11.67	2.00	0.00	100.00	0.00	0.00	0.00	98.00	2.00	0.00
24	1.00	3.00	2.00	3.00	0.00	84.00	12.00	4.00	0.00	100.00	0.00	0.00	0.00	99.67	0.33	0.00
25	1.00	3.00	3.00	1.00	0.00	17.00	47.67	35.33	0.00	82.00	16.67	1.33	0.00	25.33	52.67	22.00
26	1.00	3.00	3.00	2.00	0.00	1.67	62.00	36.33	0.00	41.67	58.33	0.00	0.00	4.33	85.00	10.67
27	1.00	3.00	3.00	3.00	0.00	0.00	68.00	32.00	0.00	3.33	96.67	0.00	0.00	0.00	94.00	6.00
28	2.00	1.00	1.00	1.00	0.00	77.33	18.67	4.00	3.33	96.67	0.00	0.00	0.00	90.67	8.67	0.67
29	2.00	1.00	1.00	2.00	0.00	80.67	14.67	4.67	0.00	100.00	0.00	0.00	0.00	97.33	2.33	0.33
30	2.00	1.00	1.00	3.00	0.00	75.00	20.00	5.00	0.00	100.00	0.00	0.00	0.00	99.00	1.00	0.00
31	2.00	1.00	2.00	1.00	0.00	70.00	24.33	5.67	0.00	100.00	0.00	0.00	0.00	85.33	12.33	2.33
32	2.00	1.00	2.00	2.00	0.00	56.33	33.67	10.00	0.00	100.00	0.00	0.00	0.00	89.67	8.67	1.67
33	2.00	1.00	2.00	3.00	0.00	43.00	44.67	12.33	0.00	100.00	0.00	0.00	0.00	90.33	9.00	0.67
34	2.00	1.00	3.00	1.00	0.00	0.00	59.33	40.67	0.00	0.33	99.00	0.67	0.00	0.00	75.00	25.00
35	2.00	1.00	3.00	2.00	0.00	0.00	65.33	34.67	0.00	0.00	99.33	0.67	0.00	0.00	88.67	11.33
36	2.00	1.00	3.00	3.00	0.00	0.00	68.67	31.33	0.00	0.00	100.00	0.00	0.00	0.00	97.00	3.00
37	2.00	2.00	1.00	1.00	0.33	74.33	23.33	2.00	11.67	88.00	0.33	0.00	0.67	88.67	10.00	0.67
38	2.00	2.00	1.00	2.00	0.00	72.67	22.00	5.33	0.00	100.00	0.00	0.00	0.00	96.00	3.33	0.67
39	2.00	2.00	1.00	3.00	0.00	65.67	29.33	5.00	0.00	100.00	0.00	0.00	0.00	98.00	2.00	0.00
40	2.00	2.00	2.00	1.00	0.00	70.67	20.33	9.00	0.00	100.00	0.00	0.00	0.00	88.33	8.67	3.00
41	2.00	2.00	2.00	2.00	0.00	50.33	39.00	10.67	0.00	99.67	0.33	0.00	0.00	87.33	11.67	1.00
42	2.00	2.00	2.00	3.00	0.00	36.67	50.33	13.00	0.00	100.00	0.00	0.00	0.00	83.33	15.67	1.00
43	2.00	2.00	3.00	1.00	0.00	0.00	63.00	37.00	0.00	6.67	93.00	0.33	0.00	0.00	75.67	24.33
44	2.00	2.00	3.00	2.00	0.00	0.00	67.33	32.67	0.00	0.00	100.00	0.00	0.00	0.00	92.67	7.33
45	2.00	2.00	3.00	3.00	0.00	0.00	69.00	31.00	0.00	0.00	100.00	0.00	0.00	0.00	95.67	4.33
46	2.00	3.00	1.00	1.00	0.00	74.67	23.00	2.33	6.67	93.33	0.00	0.00	0.33	87.33	12.00	0.33
47	2.00	3.00	1.00	2.00	0.00	73.33	19.67	7.00	0.00	100.00	0.00	0.00	0.00	96.67	3.00	0.33
48	2.00	3.00	1.00	3.00	0.00	64.33	31.67	4.00	0.00	100.00	0.00	0.00	0.00	96.67	3.33	0.00
49	2.00	3.00	2.00	1.00	0.00	77.33	17.00	5.67	0.00	100.00	0.00	0.00	0.00	87.33	10.00	2.67
50	2.00	3.00	2.00	2.00	0.00	73.00	22.00	5.00	0.00	100.00	0.00	0.00	0.00	96.00	3.33	0.67
51	2.00	3.00	2.00	3.00	0.00	61.33	30.33	8.33	0.00	100.00	0.00	0.00	0.00	94.67	4.67	0.67
52	2.00	3.00	3.00	1.00	0.00	13.00	47.33	39.67	0.00	78.00	21.00	1.00	0.00	22.00	52.67	25.33
53	2.00	3.00	3.00	2.00	0.00	0.33	65.67	34.00	0.00	35.67	64.00	0.33	0.00	3.33	85.00	11.67
54	2.00	3.00	3.00	3.00	0.00	0.00	71.67	28.33	0.00	3.00	96.67	0.33	0.00	0.00	94.33	5.67
55	3.00	1.00	1.00	1.00	0.00	76.33	19.00	4.67	2.67	97.33	0.00	0.00	0.00	88.67	9.67	1.67
56	3.00	1.00	1.00	2.00	0.00	78.33	15.00	6.67	0.00	100.00	0.00	0.00	0.00	97.33	2.00	0.67
57	3.00	1.00	1.00	3.00	0.00	74.00	21.33	4.67	0.00	100.00	0.00	0.00	0.00	99.00	1.00	0.00
58	3.00	1.00	2.00	1.00	0.00	67.33	27.00	5.67	0.00	99.67	0.00	0.33	0.00	84.33	14.00	1.67
59	3.00	1.00	2.00	2.00	0.00	59.67	35.33	5.00	0.00	100.00	0.00	0.00	0.00	93.00	6.33	0.67
60	3.00	1.00	2.00	3.00	0.00	53.67	38.67	7.67	0.00	100.00	0.00	0.00	0.00	92.00	7.33	0.67
61	3.00	1.00	3.00	1.00	0.00	0.00	57.67	42.33	0.00	3.67	93.67	2.67	0.00	0.00	70.67	29.33
62	3.00	1.00	3.00	2.00	0.00	0.00	60.33	39.67	0.00	0.00	98.67	1.33	0.00	0.00	90.00	10.00
63	3.00	1.00	3.00	3.00	0.00	0.00	67.00	33.00	0.00	0.00	100.00	0.00	0.00	0.00	96.00	4.00
64	3.00	2.00	1.00	1.00	0.00	75.00	23.33	1.67	10.33	89.67	0.00	0.00	0.67	87.00	11.67	0.67
65	3.00	2.00	1.00	2.00	0.00	75.00	21.00	4.00	0.00	100.00	0.00	0.00	0.00	94.67	5.00	0.33
66	3.00	2.00	1.00	3.00	0.00	65.33	30.00	4.67	0.00	100.00	0.00	0.00	0.00	98.67	1.33	0.00
67	3.00	2.00	2.00	1.00	0.00	71.33	23.67	5.00	0.00	100.00	0.00	0.00	0.00	89.00	9.00	2.00
68	3.00	2.00	2.00	2.00	0.00	58.67	35.00	6.33	0.00	99.33	0.67	0.00	0.00	90.67	8.67	0.67
69	3.00	2.00	2.00	3.00	0.00	46.33	46.00	7.67	0.00	100.00	0.00	0.00	0.00	88.33	11.67	0.00
70	3.00	2.00	3.00	1.00	0.00	0.00	52.33	47.67	0.00	15.67	80.00	4.33	0.00	0.00	66.00	34.00
71	3.00	2.00	3.00	2.00	0.00	0.00	61.67	38.33	0.00	0.00	97.67	2.33	0.00	0.00	87.00	13.00
72	3.00	2.00	3.00	3.00	0.00	0.00	66.33	33.67	0.00	0.00	99.67	0.33	0.00	0.00	97.33	2.67
73	3.00	3.00	1.00	1.00	0.00	74.67	22.67	2.67	6.67	93.33	0.00	0.00	0.00	88.67	11.00	0.33
74	3.00	3.00	1.00	2.00	0.00	69.23	23.75	7.02	0.00	100.00	0.00	0.00	0.00	95.32	4.35	0.33
75	3.00	3.00	1.00	3.00	0.00	61.67	33.00	5.33	0.00	100.00	0.00	0.00	0.00	97.67	2.33	0.00
76	3.00	3.00	2.00	1.00	0.00	80.00	16.67	3.33								

Tabelle 42: (Fortsetzung) Prozentuale Klassenidentifikation nach allen Konfigurationsparametern (Referenz).

	invar	mixing	sep	nobs	1.aic	2.aic	3.aic	4.aic	1.bic	2.bic	3.bic	4.bic	1.abic	2.abic	3.abic	4.abic
83	4.00	1.00	1.00	2.00	0.00	79.93	15.38	4.68	1.34	98.66	0.00	0.00	0.00	96.32	3.34	0.33
84	4.00	1.00	1.00	3.00	0.00	75.33	20.00	4.67	0.00	100.00	0.00	0.00	0.00	99.00	1.00	0.00
85	4.00	1.00	2.00	1.00	0.00	68.00	24.67	7.33	0.33	99.33	0.33	0.00	0.00	83.33	13.33	3.33
86	4.00	1.00	2.00	2.00	0.00	36.00	54.33	9.67	0.00	99.67	0.33	0.00	0.00	80.00	19.00	1.00
87	4.00	1.00	2.00	3.00	0.00	11.00	76.33	12.67	0.00	96.67	3.33	0.00	0.00	53.00	45.00	2.00
88	4.00	1.00	3.00	1.00	0.00	0.00	45.67	54.33	0.00	0.00	91.67	8.33	0.00	0.00	58.00	42.00
89	4.00	1.00	3.00	2.00	0.00	0.00	37.67	62.33	0.00	0.00	87.67	12.33	0.00	0.00	61.00	39.00
90	4.00	1.00	3.00	3.00	0.00	0.00	31.00	69.00	0.00	0.00	80.00	20.00	0.00	0.00	54.00	46.00
91	4.00	2.00	1.00	1.00	3.67	77.33	16.33	2.67	66.00	34.00	0.00	0.00	9.33	83.00	7.33	0.33
92	4.00	2.00	1.00	2.00	0.33	80.67	16.00	3.00	10.67	89.33	0.00	0.00	1.00	94.00	4.33	0.67
93	4.00	2.00	1.00	3.00	0.00	72.67	21.00	6.33	0.67	99.33	0.00	0.00	0.00	99.33	0.67	0.00
94	4.00	2.00	2.00	1.00	0.00	66.33	27.00	6.67	2.00	97.67	0.33	0.00	0.00	83.00	14.33	2.67
95	4.00	2.00	2.00	2.00	0.00	35.33	54.67	10.00	0.00	98.67	1.33	0.00	0.00	76.67	22.67	0.67
96	4.00	2.00	2.00	3.00	0.00	5.00	81.67	13.33	0.00	89.67	10.33	0.00	0.00	42.00	56.33	1.67
97	4.00	2.00	3.00	1.00	0.00	0.00	52.67	47.33	0.00	1.33	94.67	4.00	0.00	0.00	66.67	33.33
98	4.00	2.00	3.00	2.00	0.00	0.00	58.33	41.67	0.00	0.00	96.33	3.67	0.00	0.00	81.33	18.67
99	4.00	2.00	3.00	3.00	0.00	0.00	60.00	40.00	0.00	0.00	95.00	5.00	0.00	0.00	81.67	18.33
100	4.00	3.00	1.00	1.00	2.33	80.00	16.00	1.67	56.33	43.67	0.00	0.00	5.33	87.33	6.67	0.67
101	4.00	3.00	1.00	2.00	0.00	81.67	13.67	4.67	6.67	93.33	0.00	0.00	1.00	98.00	1.00	0.00
102	4.00	3.00	1.00	3.00	0.00	78.33	17.67	4.00	0.33	99.67	0.00	0.00	0.00	99.33	0.67	0.00
103	4.00	3.00	2.00	1.00	0.00	52.00	39.00	9.00	0.00	99.00	1.00	0.00	0.00	69.00	25.67	5.33
104	4.00	3.00	2.00	2.00	0.00	15.38	72.58	12.04	0.00	96.66	3.34	0.00	0.00	48.49	48.49	3.01
105	4.00	3.00	2.00	3.00	0.00	1.67	89.33	9.00	0.00	68.00	32.00	0.00	0.00	18.33	80.00	1.67
106	4.00	3.00	3.00	1.00	0.00	2.33	66.67	31.00	0.00	48.67	50.33	1.00	0.00	6.67	71.67	21.67
107	4.00	3.00	3.00	2.00	0.00	0.00	78.33	21.67	0.00	5.67	94.00	0.33	0.00	0.67	93.67	5.67
108	4.00	3.00	3.00	3.00	0.00	0.00	79.67	20.33	0.00	0.00	100.00	0.00	0.00	0.00	96.67	3.33
109	5.00	1.00	1.00	1.00	0.00	61.67	28.00	10.33	9.00	90.67	0.33	0.00	0.33	79.67	16.00	4.00
110	5.00	1.00	1.00	2.00	0.00	40.67	40.33	19.00	0.00	99.00	1.00	0.00	0.00	80.00	17.67	2.33
111	5.00	1.00	1.00	3.00	0.00	23.67	40.67	35.67	0.00	98.67	1.00	0.33	0.00	75.00	19.00	6.00
112	5.00	1.00	2.00	1.00	0.00	37.67	40.67	21.67	0.00	97.67	2.33	0.00	0.00	53.00	33.67	13.33
113	5.00	1.00	2.00	2.00	0.00	6.67	46.33	47.00	0.00	90.33	9.67	0.00	0.00	35.67	47.00	17.33
114	5.00	1.00	2.00	3.00	0.00	0.00	28.67	71.33	0.00	56.67	39.33	4.00	0.00	10.00	52.67	37.33
115	5.00	1.00	3.00	1.00	0.00	22.00	40.67	37.33	0.00	91.33	7.33	1.33	0.00	35.33	37.67	27.00
116	5.00	1.00	3.00	2.00	0.00	1.00	30.00	69.00	0.00	78.33	16.67	5.00	0.00	13.00	42.33	44.67
117	5.00	1.00	3.00	3.00	0.00	0.00	8.33	91.67	0.00	32.00	41.67	26.33	0.00	2.33	26.67	71.00
118	5.00	2.00	1.00	1.00	1.00	67.00	24.33	7.67	34.33	65.67	0.00	0.00	2.67	78.33	16.33	2.67
119	5.00	2.00	1.00	2.00	0.00	42.67	39.00	18.33	2.00	97.33	0.67	0.00	0.00	80.00	18.00	2.00
120	5.00	2.00	1.00	3.00	0.00	21.67	47.67	30.67	0.00	98.00	2.00	0.00	0.00	73.33	24.33	2.33
121	5.00	2.00	2.00	1.00	0.00	47.67	36.00	16.33	4.67	94.67	0.67	0.00	0.00	63.33	27.67	9.00
122	5.00	2.00	2.00	2.00	0.00	14.00	50.00	36.00	0.00	91.67	8.00	0.33	0.00	45.00	42.67	12.33
123	5.00	2.00	2.00	3.00	0.00	1.33	40.33	58.33	0.00	68.67	29.33	2.00	0.00	20.00	54.67	25.33
124	5.00	2.00	3.00	1.00	0.00	8.33	42.67	49.00	0.00	89.67	9.00	1.33	0.00	19.67	41.33	39.00
125	5.00	2.00	3.00	2.00	0.00	0.00	20.33	79.67	0.00	53.33	31.00	15.67	0.00	4.00	36.00	60.00
126	5.00	2.00	3.00	3.00	0.00	0.00	7.67	92.33	0.00	10.00	34.33	55.67	0.00	0.33	16.33	83.33
127	5.00	3.00	1.00	1.00	0.33	69.67	25.67	4.33	32.67	66.67	0.67	0.00	1.00	83.33	14.00	1.67
128	5.00	3.00	1.00	2.00	0.00	41.00	45.00	14.00	2.00	96.33	1.67	0.00	0.00	77.67	19.67	2.67
129	5.00	3.00	1.00	3.00	0.00	16.67	47.00	36.33	0.00	95.33	4.67	0.00	0.00	64.33	29.00	6.67
130	5.00	3.00	2.00	1.00	0.00	46.33	38.67	15.00	1.33	97.67	1.00	0.00	0.00	68.67	24.00	7.33
131	5.00	3.00	2.00	2.00	0.00	16.33	45.33	38.33	0.00	94.00	5.33	0.67	0.00	55.00	30.00	15.00
132	5.00	3.00	2.00	3.00	0.00	2.67	34.00	63.33	0.00	77.00	18.33	4.67	0.00	25.33	37.33	37.33
133	5.00	3.00	3.00	1.00	0.00	0.67	41.14	58.19	0.00	47.16	45.82	7.02	0.00	2.68	51.51	45.82
134	5.00	3.00	3.00	2.00	0.00	0.00	18.00	82.00	0.00	2.67	62.33	35.00	0.00	0.00	33.00	67.00
135	5.00	3.00	3.00	3.00	0.00	0.00	5.00	95.00	0.00	0.00	42.00	58.00	0.00	0.00	17.33	82.67
136	6.00	1.00	1.00	1.00	0.00	60.67	29.00	10.33	7.00	93.00	0.00	0.00	0.00	79.00	16.67	4.33
137	6.00	1.00	1.00	2.00	0.00	41.33	35.67	23.00	0.00	99.00	0.67	0.33	0.00	77.00	19.33	3.67
138	6.00	1.00	1.00	3.00	0.00	20.33	41.00	38.67	0.00	97.00	3.00	0.00	0.00	71.67	19.33	9.00
139	6.00	1.00	2.00	1.00	0.00	46.67	35.33	18.00	1.33	97.67	1.00	0.00	0.00	64.67	24.67	10.67
140	6.00	1.00	2.00	2.00	0.00	15.00	39.67	45.33	0.00	95.33	4.67	0.00	0.00	52.33	29.67	18.00
141	6.00	1.00	2.00	3.00	0.00	3.67	37.33	59.00	0.00	77.67	20.33	2.00	0.00	23.67	45.00	31.33
142	6.00	1.00	3.00	1.00	0.00	49.00	31.00	20.00	0.00	98.33	1.67	0.00	0.00	70.33	19.67	10.00
143	6.00	1.00	3.00	2.00	0.00	30.33	33.00	36.67	0.00	97.00	2.67	0.33	0.00	63.33	25.00	11.67
144	6.00	1.00	3.00	3.00	0.00	7.00	28.67	64.33	0.00	90.67	8.00	1.33	0.00	50.33	30.33	19.33
145	6.00	2.00	1.00	1.00	1.00	61.00	27.67	10.33	34.67	65.33	0.00	0.00	2.00	78.67	14.33	5.00
146	6.00	2.00	1.00	2.00	0.00	43.67	34.67	21.67	1.67	96.67	1.33	0.33	0.00	80.00	17.00	3.00
147	6.00	2.00	1.00	3.00	0.00	22.33	45.00	32.67	0.00	97.00	2.67	0.33	0.00	70.00	22.00	8.00
148	6.00	2.00	2.00	1.00	0.00	49.33	36.67	14.00	11.33	88.33	0.33	0.00	0.00	66.33	25.33	8.33
149	6.00	2.00	2.00	2.00	0.00	15.00	54.00	31.00	0.00	91.00	8.67	0.33	0.00	52.33	38.00	9.67
150	6.00	2.00	2.00	3.00	0.00	2.33	46.33	51.33	0.00	73.33	25.00	1.67	0.00	23.67	52.33	24.00
151	6.00	2.00	3.00	1.00	0.00	58.00	25.00	17.00	0.00	99.00	1.00	0.00	0.00	72.67	16.33	11.00
152	6.00	2.00	3.00	2.00	0.00	33.11	21.40	45.48	0.00	97.99	1.67	0.33	0.00	72.58	13.71	13.71
153	6.00	2.00	3.00	3.00	0.00	7.33	17.33	75.33	0.00	95.00	1.67	3.33	0.00	53.00	15.00	32.00
154	6.00	3.00	1.00	1.00	0.33	63.67	28.33	7.67	39.67	60.00	0.33	0.00	2.00	76.00	17.00	5.00
155	6.00	3.00	1.00	2.00	0.00	31.00	43.00	26.00	2.33	96.00	1.67	0.00	0.00	70.33	24.33	5.33
156	6.00	3.00	1.00	3.00	0.00	8.00	52.67	39.33	0.00	91.67	8.33	0.00	0.00	50.67	36.33	13.00
157	6.00	3.00	2.00	1.00	0.00	56.33	35.33	8.33	4.67	94.00	1.33	0.00				

Tabelle 43: Logitkoeffizienten der Modelle zur Klassenidentifikation nach AIC, BIC, aBIC

	<i>Dependent variable:</i>		
	Korrekte KI nach AIC	Korrekte KI nach BIC	Korrekte KI nach aBIC
	(1)	(2)	(3)
nobs2	0.011 (0.068)	0.193 (0.128)	0.165** (0.079)
nobs3	0.110 (0.068)	0.726*** (0.128)	0.196** (0.079)
mixing2	−0.185*** (0.068)	−0.450*** (0.129)	−0.127 (0.078)
mixing3	−0.286*** (0.069)	−3.022*** (0.130)	−0.349*** (0.080)
invar2	0.005 (0.075)	0.081 (0.142)	0.084 (0.088)
invar3	0.030 (0.076)	−0.039 (0.145)	0.095 (0.089)
invar4	−0.274*** (0.076)	−0.601*** (0.140)	−0.410*** (0.089)
invar5	0.132* (0.075)	−2.986*** (0.141)	0.134 (0.088)
invar6	−0.037 (0.076)	−4.830*** (0.218)	−0.526*** (0.091)
sep2	0.479*** (0.025)	2.125*** (0.113)	0.951*** (0.033)
sep3	0.942*** (0.024)	6.751*** (0.111)	2.777*** (0.032)
nobs2:mixing2	0.120** (0.058)	0.544*** (0.116)	0.117* (0.068)
nobs3:mixing2	0.203*** (0.058)	0.327*** (0.114)	0.177*** (0.068)
nobs2:mixing3	0.087 (0.059)	1.325*** (0.115)	0.130* (0.070)
nobs3:mixing3	0.143** (0.059)	1.841*** (0.116)	0.194*** (0.069)
nobs2:invar2	0.185** (0.083)	−0.065 (0.151)	−0.016 (0.096)
nobs3:invar2	0.354*** (0.082)	−0.157 (0.154)	0.070 (0.096)
nobs2:invar3	0.107 (0.085)	0.027 (0.159)	−0.076 (0.099)
nobs3:invar3	0.245*** (0.084)	−0.226 (0.160)	−0.012 (0.098)
nobs2:invar4	0.382*** (0.083)	−0.064 (0.151)	0.154 (0.096)
nobs3:invar4	0.594*** (0.082)	−0.019 (0.151)	0.588*** (0.095)
nobs2:invar5	0.002 (0.083)	0.223 (0.150)	−0.079 (0.096)
nobs3:invar5	−0.535*** (0.084)	0.677*** (0.152)	−0.212** (0.096)
nobs2:invar6	0.120 (0.084)	0.463* (0.238)	0.004 (0.100)
nobs3:invar6	−0.020 (0.083)	1.211*** (0.225)	0.327*** (0.099)
mixing2:invar2	0.240*** (0.082)	0.076 (0.157)	0.110 (0.094)
mixing3:invar2	0.103 (0.083)	0.146 (0.151)	0.005 (0.096)
mixing2:invar3	0.162** (0.082)	−0.097 (0.157)	0.032 (0.095)
mixing3:invar3	−0.466*** (0.085)	−1.292*** (0.157)	−0.941*** (0.101)
mixing2:invar4	0.387*** (0.082)	0.791*** (0.156)	0.495*** (0.095)
mixing3:invar4	0.926*** (0.082)	2.330*** (0.152)	1.261*** (0.095)
mixing2:invar5	0.103 (0.083)	0.067 (0.149)	−0.076 (0.095)
mixing3:invar5	0.192** (0.083)	2.855*** (0.150)	−0.014 (0.096)
mixing2:invar6	0.062 (0.083)	0.115 (0.181)	−0.095 (0.098)
mixing3:invar6	0.200** (0.083)	1.952*** (0.179)	0.228** (0.098)
Constant	−1.174*** (0.059)	−4.851*** (0.150)	−2.331*** (0.071)
Observations	48,594	48,594	48,594
Log Likelihood	−30,478.880	−10,075.780	−23,644.060
Akaike Inf. Crit.	61,029.750	20,223.550	47,360.130

Note:

*p<0.1; **p<0.05; ***p<0.01

Tabelle 44: Odds-Ratios der Modelle zur Klassenidentifikation nach AIC, BIC, aBIC

	<i>Dependent variable:</i>		
	Korrekte KI nach AIC	Korrekte KI nach BIC	Korrekte KI nach aBIC
	(1)	(2)	(3)
nobs2	1.011	1.213	1.179**
nobs3	1.117	2.066***	1.216**
mixing2	0.831***	0.638***	0.881
mixing3	0.751***	0.049***	0.705***
invar2	1.005	1.084	1.088
invar3	1.030	0.962	1.099
invar4	0.761***	0.548***	0.664***
invar5	1.141*	0.050***	1.143
invar6	0.964	0.008***	0.591***
sep2	1.614***	8.375***	2.587***
sep3	2.564***	855.031***	16.064***
nobs2:mixing2	1.128**	1.722***	1.124*
nobs3:mixing2	1.225***	1.387***	1.193***
nobs2:mixing3	1.090	3.761***	1.139*
nobs3:mixing3	1.154**	6.301***	1.214***
nobs2:invar2	1.204**	0.937	0.984
nobs3:invar2	1.424***	0.855	1.072
nobs2:invar3	1.113	1.027	0.927
nobs3:invar3	1.277***	0.798	0.988
nobs2:invar4	1.466***	0.938	1.167
nobs3:invar4	1.811***	0.981	1.800***
nobs2:invar5	1.002	1.250	0.924
nobs3:invar5	0.586***	1.967***	0.809**
nobs2:invar6	1.128	1.589*	1.004
nobs3:invar6	0.980	3.356***	1.387***
mixing2:invar2	1.271***	1.079	1.117
mixing3:invar2	1.108	1.157	1.005
mixing2:invar3	1.176**	0.908	1.032
mixing3:invar3	0.628***	0.275***	0.390***
mixing2:invar4	1.472***	2.206***	1.640***
mixing3:invar4	2.523***	10.278***	3.528***
mixing2:invar5	1.108	1.069	0.927
mixing3:invar5	1.212**	17.381***	0.986
mixing2:invar6	1.064	1.121	0.909
mixing3:invar6	1.221**	7.045***	1.257**
Constant	0.309***	0.008***	0.097***
Observations	48,594	48,594	48,594
Log Likelihood	-30,478.880	-10,075.780	-23,644.060
Akaike Inf. Crit.	61,029.750	20,223.550	47,360.130

Note:

*p<0.1; **p<0.05; ***p<0.01

Tabelle 45: Logitkoeffizienten der Modelle zur Klassenidentifikation nach BLRT

	<i>Dependent variable:</i>
	Korrekte KI nach BLRT
nobs2	0.923*** (0.084)
nobs3	1.065*** (0.083)
mixing2	0.137*** (0.035)
mixing3	−0.187*** (0.036)
invar2	0.083 (0.183)
invar3	0.414** (0.171)
invar4	0.274 (0.176)
invar5	2.636*** (0.142)
invar6	2.606*** (0.142)
sep2	2.428*** (0.161)
sep3	6.430*** (0.170)
invar2:sep2	0.180 (0.198)
invar3:sep2	−0.223 (0.188)
invar4:sep2	0.128 (0.192)
invar5:sep2	−2.365*** (0.161)
invar6:sep2	−2.441*** (0.162)
invar2:sep3	−0.665*** (0.208)
invar3:sep3	−1.020*** (0.198)
invar4:sep3	−1.818*** (0.198)
invar5:sep3	−5.792*** (0.170)
invar6:sep3	−5.995*** (0.171)
nobs2:sep2	−0.267*** (0.101)
nobs3:sep2	−0.312*** (0.100)
nobs2:sep3	−0.937*** (0.103)
nobs3:sep3	−1.383*** (0.102)
Constant	−4.114*** (0.148)
Observations	32,399
Log Likelihood	−14,923.330
Akaike Inf. Crit.	29,898.670
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Tabelle 46: Odds-Ratios der Modelle zur Klassenidentifikation nach BLRT

<i>Dependent variable:</i>	
Korrekte KI nach BLRT	
nobs2	2.518***
nobs3	2.900***
mixing2	1.147***
mixing3	0.829***
invar2	1.087
invar3	1.513**
invar4	1.316
invar5	13.957***
invar6	13.546***
sep2	11.332***
sep3	620.163***
invar2:sep2	1.198
invar3:sep2	0.801
invar4:sep2	1.137
invar5:sep2	0.094***
invar6:sep2	0.087***
invar2:sep3	0.514***
invar3:sep3	0.361***
invar4:sep3	0.162***
invar5:sep3	0.003***
invar6:sep3	0.002***
nobs2:sep2	0.766***
nobs3:sep2	0.732***
nobs2:sep3	0.392***
nobs3:sep3	0.251***
Constant	0.016***
Observations	32,399
Log Likelihood	−14,923.330
Akaike Inf. Crit.	29,898.670
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

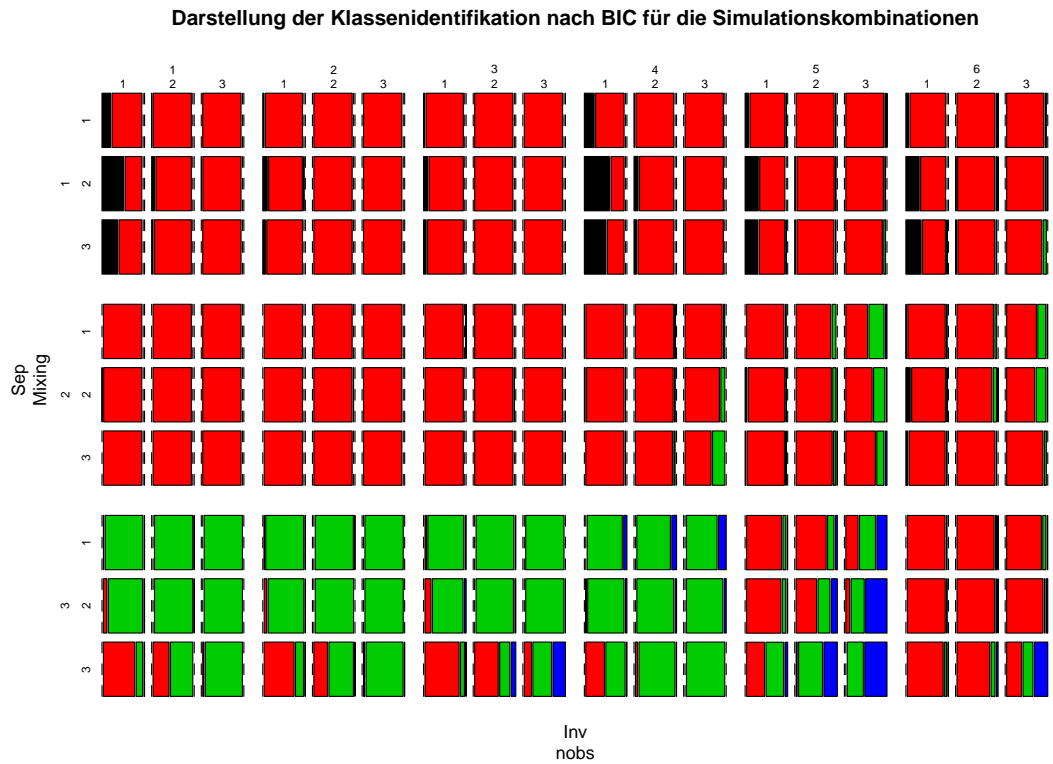
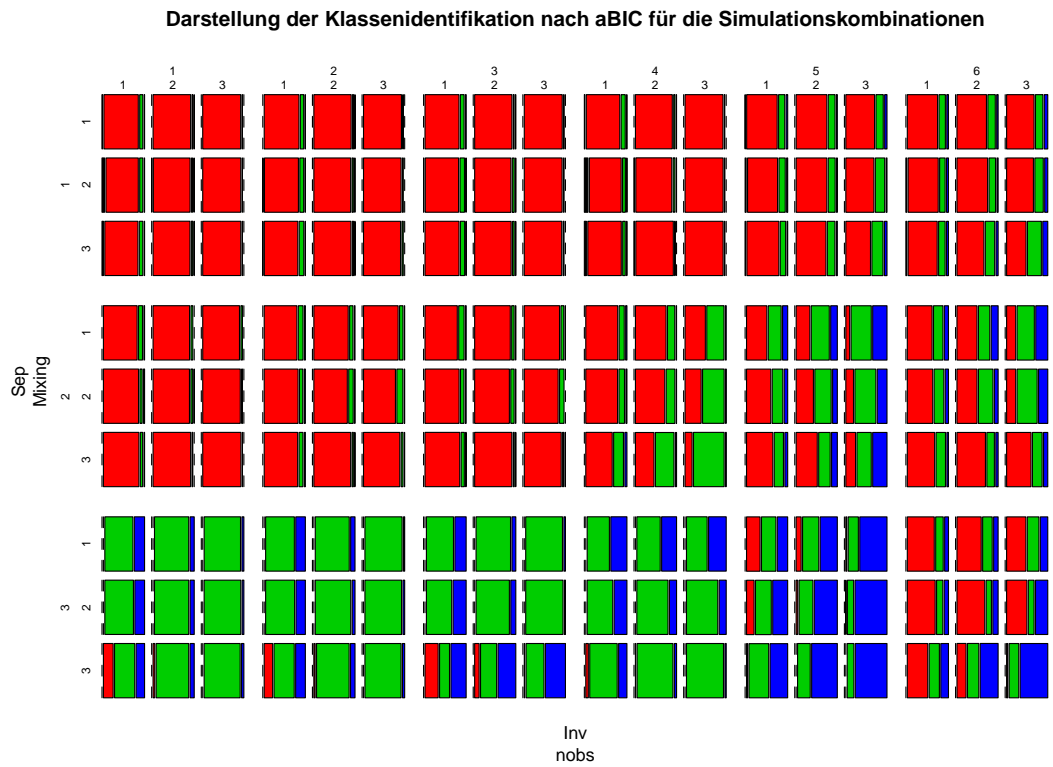


Abbildung 13: Mosaik-Plots zur Klassenidentifikation (aBIC, BIC)

Legende: Schwarz=1-Klasse, Rot=2-Klassen, Grün= 3-Klassen; Blau=4-Klassen.
Replikationsanzahl pro Konfiguration jeweils 300/299.

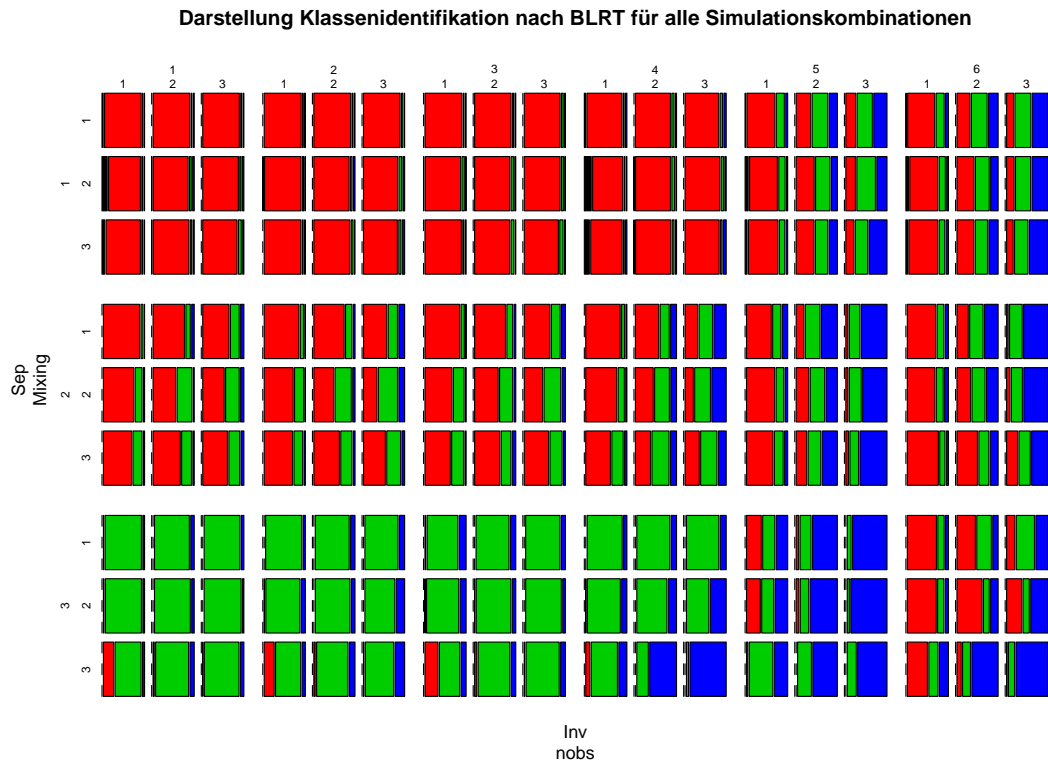
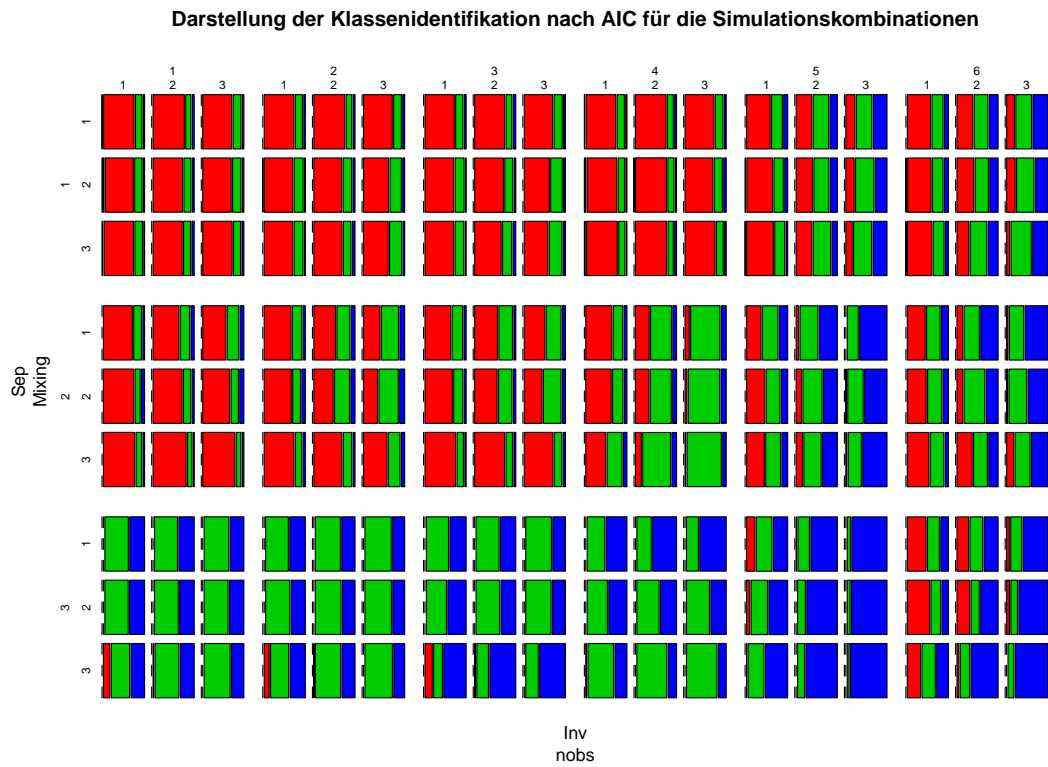


Abbildung 14: Mosaik-Plots zur Klassenidentifikation (AIC, BLRT)

Legende: Schwarz=1-Klasse, Rot=2-Klassen, Grün= 3-Klassen; Blau=4-Klassen.
Replikationsanzahlen pro Konfiguration für AIC: 300/299; BLRT:200/199.

A.2 Relativer Bias Referenz

Tabelle 47: Referenztable: Relativer Bias für alle Simulationskonfigurationen (Teil 1)

Confignum	Invar	Sep	Nobs	Mixing	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(S_C1)	M(S_C2)	M(S_C3)	N
1	1.000	1.000	1.000	1.000	-0.361	-0.309	0.564	1.374	0.190	-0.321	-0.142	1.370	52
2	1.000	2.000	1.000	1.000	-0.140	-0.182	0.718	0.481	0.063	-0.109	-0.061	1.629	61
3	1.000	3.000	1.000	1.000	-0.014	-0.029	-0.030	0.004	-0.006	0.001	0.002	0.023	297
4	2.000	1.000	1.000	1.000	-0.318	-0.293	0.483	1.956	0.395	-0.306	-0.310	1.151	57
5	2.000	2.000	1.000	1.000	-0.093	-0.096	0.866	3.876	0.257	-0.137	-0.087	1.144	75
6	2.000	3.000	1.000	1.000	0.022	-0.001	0.033	0.053	0.073	-0.009	-0.033	-0.447	298
7	3.000	1.000	1.000	1.000	-0.220	-0.444	0.306	2.096	0.491	-0.402	-0.208	0.537	58
8	3.000	2.000	1.000	1.000	-0.010	-0.230	1.041	0.911	0.335	-0.256	-0.062	1.188	82
9	3.000	3.000	1.000	1.000	0.131	-0.117	0.083	0.148	0.123	-0.034	-0.053	-0.425	284
10	4.000	1.000	1.000	1.000	-0.382	-0.330	-0.044	1.705	0.045	-0.360	-0.058	1.470	57
11	4.000	2.000	1.000	1.000	-0.078	-0.283	-0.125	0.166	-0.279	-0.383	0.181	2.265	75
12	4.000	3.000	1.000	1.000	0.021	0.009	-0.021	-0.259	0.314	-0.378	0.769	0.027	275
13	5.000	1.000	1.000	1.000	-0.433	-0.357	-0.419	2.699	1.658	0.017	-0.488	0.681	90
14	5.000	2.000	1.000	1.000	-0.266	-0.107	-0.162	1.467	-0.043	-0.225	-0.278	1.430	133
15	5.000	3.000	1.000	1.000	0.087	-0.015	0.050	0.421	0.054	-0.361	0.379	0.064	146
16	6.000	1.000	1.000	1.000	-0.354	-0.612	-0.539	2.739	1.477	-0.239	-0.448	-0.537	92
17	6.000	2.000	1.000	1.000	-0.204	-0.301	0.189	3.610	0.167	-0.344	-0.424	0.442	113
18	6.000	3.000	1.000	1.000	0.174	0.145	0.654	0.743	0.064	-0.292	0.287	0.505	102
19	1.000	1.000	1.000	2.000	-0.346	-0.304	0.216	1.097	-0.178	-0.198	-0.189	0.776	58
20	1.000	2.000	1.000	2.000	-0.174	-0.176	0.450	0.486	0.066	-0.010	-0.110	0.394	40
21	1.000	3.000	1.000	2.000	-0.016	-0.032	-0.039	0.004	-0.008	0.001	0.001	0.007	285
22	2.000	1.000	1.000	2.000	-0.247	-0.353	0.440	1.437	0.119	-0.433	-0.341	0.479	71
23	2.000	2.000	1.000	2.000	-0.107	-0.097	0.643	0.841	0.233	-0.091	0.098	0.014	62
24	2.000	3.000	1.000	2.000	0.014	-0.014	0.021	0.060	0.088	-0.011	-0.037	-0.472	292
25	3.000	1.000	1.000	2.000	-0.179	-0.513	0.441	1.499	0.242	-0.400	-0.404	0.214	70
26	3.000	2.000	1.000	2.000	0.069	-0.371	0.914	0.700	0.287	-0.248	0.060	0.076	74
27	3.000	3.000	1.000	2.000	0.114	-0.136	0.082	0.161	0.175	-0.041	-0.057	-0.433	260
28	4.000	1.000	1.000	2.000	-0.332	-0.316	0.490	1.266	-0.145	-0.536	-0.194	2.520	49
29	4.000	2.000	1.000	2.000	-0.003	-0.313	0.131	0.227	-0.303	-0.187	-0.025	2.170	83
30	4.000	3.000	1.000	2.000	0.049	0.023	0.037	-0.235	0.351	-0.358	0.761	0.025	288
31	5.000	1.000	1.000	2.000	-0.394	-0.285	-0.182	3.007	1.175	-0.007	-0.333	0.409	77
32	5.000	2.000	1.000	2.000	-0.212	0.079	-0.273	1.621	-0.117	-0.216	-0.218	1.278	115
33	5.000	3.000	1.000	2.000	0.181	0.105	-0.034	0.444	0.181	-0.368	0.402	0.042	153
34	6.000	1.000	1.000	2.000	-0.328	-0.588	-0.549	2.323	0.955	-0.170	-0.597	-0.553	87
35	6.000	2.000	1.000	2.000	-0.144	-0.316	-0.183	1.447	-0.068	-0.313	-0.459	-0.257	115
36	6.000	3.000	1.000	2.000	0.304	0.305	0.929	0.833	0.255	-0.216	0.189	0.152	77
37	1.000	1.000	1.000	3.000	-0.323	-0.264	0.322	1.403	0.038	-0.125	-0.308	0.787	59
38	1.000	2.000	1.000	3.000	-0.167	-0.292	0.247	0.493	0.081	-0.091	0.365	0.442	44
39	1.000	3.000	1.000	3.000	-0.019	-0.074	-0.141	0.018	0.020	-0.002	0.009	0.067	182
40	2.000	1.000	1.000	3.000	-0.305	-0.246	0.633	1.413	0.286	-0.365	-0.376	0.538	69
41	2.000	2.000	1.000	3.000	-0.100	-0.223	0.484	0.816	0.248	0.043	0.103	0.116	51
42	2.000	3.000	1.000	3.000	0.009	-0.048	-0.051	0.072	0.141	-0.018	-0.021	-0.461	189
43	3.000	1.000	1.000	3.000	-0.192	-0.442	0.579	1.606	0.304	-0.404	-0.232	0.488	69
44	3.000	2.000	1.000	3.000	-0.020	-0.392	0.494	0.990	0.393	-0.238	0.315	-0.346	51
45	3.000	3.000	1.000	3.000	0.097	-0.179	0.047	0.177	0.361	-0.078	0.060	-0.438	102
46	4.000	1.000	1.000	3.000	-0.375	-0.235	-0.007	1.465	0.096	-0.504	-0.013	1.554	48
47	4.000	2.000	1.000	3.000	-0.008	-0.269	0.236	0.255	-0.217	-0.071	0.118	1.851	120
48	4.000	3.000	1.000	3.000	0.047	0.023	0.084	-0.106	0.399	-0.280	0.634	0.096	250
49	5.000	1.000	1.000	3.000	-0.278	-0.277	-0.222	3.129	1.051	-0.048	-0.544	0.247	78
50	5.000	2.000	1.000	3.000	-0.052	-0.022	-0.046	1.493	-0.038	-0.337	-0.132	0.760	123
51	5.000	3.000	1.000	3.000	0.198	0.148	-0.289	0.106	0.334	-0.340	0.141	-0.219	205
52	6.000	1.000	1.000	3.000	-0.278	-0.550	-0.339	2.744	0.992	-0.070	-0.476	-0.930	85
53	6.000	2.000	1.000	3.000	0.040	-0.295	0.579	1.817	0.019	-0.233	-0.346	0.865	110
54	6.000	3.000	1.000	3.000	0.375	0.412	0.740	0.207	0.407	-0.230	-0.092	-0.004	111
55	1.000	1.000	2.000	1.000	-0.325	-0.155	0.486	1.258	0.044	-0.249	-0.068	1.022	40
56	1.000	2.000	2.000	1.000	-0.090	-0.116	0.559	0.368	0.063	-0.058	0.187	0.394	74
57	1.000	3.000	2.000	1.000	-0.002	-0.000	0.008	0.001	-0.004	0.001	0.002	0.014	299
58	2.000	1.000	2.000	1.000	-0.274	-0.114	0.605	1.647	0.367	-0.462	-0.194	0.922	44
59	2.000	2.000	2.000	1.000	-0.018	-0.065	0.654	0.600	0.234	-0.055	0.098	0.057	104
60	2.000	3.000	2.000	1.000	0.034	0.028	0.070	0.051	0.075	-0.009	-0.033	-0.456	298
61	3.000	1.000	2.000	1.000	-0.058	-0.439	0.315	1.689	0.441	-0.345	-0.227	0.311	45
62	3.000	2.000	2.000	1.000	0.119	-0.243	0.751	0.698	0.250	-0.219	0.183	-0.074	106
63	3.000	3.000	2.000	1.000	0.146	-0.082	0.142	0.141	0.129	-0.035	-0.056	-0.426	296
64	4.000	1.000	2.000	1.000	-0.247	-0.315	0.030	0.927	-0.148	-0.361	-0.426	1.722	47
65	4.000	2.000	2.000	1.000	-0.051	-0.151	-0.107	0.224	-0.399	-0.398	0.392	1.188	170
66	4.000	3.000	2.000	1.000	0.039	0.040	-0.023	-0.269	0.330	-0.428	0.862	0.019	263
67	5.000	1.000	2.000	1.000	-0.357	-0.287	-0.343	2.607	1.439	-0.074	-0.479	0.413	132
68	5.000	2.000	2.000	1.000	-0.175	-0.032	0.008	1.430	-0.077	-0.288	-0.282	1.261	194
69	5.000	3.000	2.000	1.000	0.113	0.082	0.101	0.451	0.073	-0.367	0.413	0.134	171
70	6.000	1.000	2.000	1.000	-0.308	-0.540	-0.471	2.634	1.621	-0.342	-0.527	-0.503	120
71	6.000	2.000	2.000	1.000	-0.106	-0.270	-0.124	1.415	-0.036	-0.351	-0.475	-0.140	143
72	6.000	3.000	2.000	1.000	0.248	0.249	0.547	0.771	0.118	-0.305	0.303	0.238	123
73	1.000	1.000	2.000	2.000	-0.302	-0.193	0.057	0.999	-0.227	-0.213	-0.180	0.334	56
74	1.000	2.000	2.000	2.000	-0.056	-0.153	0.308	0.457	-0.038	-0.048	0.187	0.291	58
75	1.000	3.000	2.000	2.000	-0.002	-0.004	-0.000	0.001	0.001	0.001	0.001	0.013	299
76	2.000	1.000	2.000	2.000	-0.213	-0.221	0.402	1.243	0.122	-0.249	-0.286	0.479	67
77	2.000	2.000	2.000	2.000	-0.016	-0.101	0.302	0.564	0.204	-0.071	0.084	-0.518	124
78	2.000	3.000	2.000	2.000	0.027	0.014	0.065	0.057	0.098	-0.012	-0.038	-0.462	300
79	3.000	1.000	2.000	2.000	-0.045	-0.410	0.134	1.508	0.305	-0.321	-0.057	-0.306	64
80	3.000	2.000	2.000	2.000	0.115	-0.253	0.307	0.690	0.262	-0.166	0.110	-0.309	106
81	3.000	3.000	2.000	2.000	0.136	-0.111	0.141	0.156	0.169	-0.042	-0.063	-0.424	293

Info: Für die 162 Simulationskonfigurationen wurden jeweils nur Replikationen berücksichtigt, welche korrekt als 3-Klassenmodell identifiziert wurden.
Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

Tabelle 48: Referenztabelle: Relativer Bias für alle Simulationskonfigurationen (Teil 2)

Confignum	Invar	Sep	Nobs	Mixing	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(S_C1)	M(S_C2)	M(S_C3)	N
82	4.000	1.000	2.000	2.000	-0.299	-0.215	-0.031	1.438	-0.125	-0.403	-0.237	1.417	48
83	4.000	2.000	2.000	2.000	0.008	-0.157	0.017	0.180	-0.329	-0.288	0.365	0.671	175
84	4.000	3.000	2.000	2.000	0.058	0.047	0.062	-0.234	0.364	-0.429	0.893	0.035	289
85	5.000	1.000	2.000	2.000	-0.279	-0.218	-0.402	2.896	1.069	0.071	-0.572	-0.059	129
86	5.000	2.000	2.000	2.000	-0.102	0.098	-0.133	1.774	-0.072	-0.192	-0.236	1.032	183
87	5.000	3.000	2.000	2.000	0.248	0.164	0.061	0.464	0.175	-0.356	0.406	0.132	161
88	6.000	1.000	2.000	2.000	-0.227	-0.490	-0.501	2.683	1.124	-0.298	-0.418	-0.857	110
89	6.000	2.000	2.000	2.000	-0.038	-0.250	-0.203	1.610	-0.133	-0.298	-0.412	-0.334	180
90	6.000	3.000	2.000	2.000	0.416	0.435	0.894	1.098	0.272	-0.237	0.379	0.378	80
91	1.000	1.000	2.000	3.000	-0.280	-0.162	0.299	1.044	-0.147	-0.084	-0.173	0.680	54
92	1.000	2.000	2.000	3.000	-0.071	-0.185	0.117	0.664	-0.015	-0.006	0.177	0.300	37
93	1.000	3.000	2.000	3.000	-0.003	-0.016	-0.016	0.001	0.011	0.001	-0.002	0.003	281
94	2.000	1.000	2.000	3.000	-0.198	-0.203	0.320	1.277	0.176	-0.221	-0.074	-0.270	61
95	2.000	2.000	2.000	3.000	-0.007	-0.150	0.295	0.863	0.349	-0.013	0.365	-0.651	68
96	2.000	3.000	2.000	3.000	0.026	0.001	0.058	0.055	0.101	-0.012	-0.041	-0.428	277
97	3.000	1.000	2.000	3.000	-0.045	-0.424	0.113	1.386	0.381	-0.385	-0.026	-0.415	72
98	3.000	2.000	2.000	3.000	0.127	-0.321	0.151	0.891	0.423	-0.198	0.337	-0.721	52
99	3.000	3.000	2.000	3.000	0.124	-0.123	0.168	0.145	0.243	-0.040	-0.016	-0.423	147
100	4.000	1.000	2.000	3.000	-0.267	-0.168	-0.319	2.271	0.039	-0.354	-0.245	1.190	41
101	4.000	2.000	2.000	3.000	0.040	-0.102	0.213	0.181	-0.222	-0.190	0.271	1.067	230
102	4.000	3.000	2.000	3.000	0.070	0.058	0.117	-0.134	0.453	-0.340	0.747	0.032	297
103	5.000	1.000	2.000	3.000	-0.205	-0.180	-0.258	5.955	0.835	-0.169	-0.520	0.159	145
104	5.000	2.000	2.000	3.000	0.016	0.060	-0.110	1.475	-0.179	-0.448	-0.056	0.573	160
105	5.000	3.000	2.000	3.000	0.232	0.188	-0.442	-0.100	0.347	-0.320	-0.052	-0.339	190
106	6.000	1.000	2.000	3.000	-0.150	-0.446	-0.408	2.859	0.814	-0.110	-0.592	-0.997	145
107	6.000	2.000	2.000	3.000	0.155	-0.127	0.145	1.982	-0.042	-0.148	-0.336	0.225	121
108	6.000	3.000	2.000	3.000	0.427	0.465	0.512	-0.081	0.382	-0.270	-0.206	-0.207	110
109	1.000	1.000	3.000	1.000	-0.231	-0.135	0.153	1.356	0.005	-0.057	0.090	0.566	62
110	1.000	2.000	3.000	1.000	-0.028	-0.014	0.523	0.439	0.060	-0.087	0.171	1.039	91
111	1.000	3.000	3.000	1.000	0.000	0.004	-0.002	-0.002	-0.004	0.000	0.000	0.006	300
112	2.000	1.000	3.000	1.000	-0.135	-0.177	0.121	1.583	0.291	-0.199	-0.053	-0.051	60
113	2.000	2.000	3.000	1.000	0.032	0.029	0.609	0.568	0.243	-0.101	0.172	0.379	141
114	2.000	3.000	3.000	1.000	0.037	0.032	0.059	0.048	0.076	-0.009	-0.035	-0.463	300
115	3.000	1.000	3.000	1.000	0.008	-0.385	0.306	1.685	0.481	-0.239	-0.205	0.286	64
116	3.000	2.000	3.000	1.000	0.181	-0.177	0.529	0.752	0.373	-0.185	0.179	0.006	116
117	3.000	3.000	3.000	1.000	0.149	-0.079	0.130	0.138	0.130	-0.036	-0.058	-0.432	300
118	4.000	1.000	3.000	1.000	-0.208	-0.191	-0.453	1.136	-0.239	-0.195	-0.312	1.223	60
119	4.000	2.000	3.000	1.000	-0.021	-0.095	-0.018	0.140	-0.464	-0.426	0.528	0.845	243
120	4.000	3.000	3.000	1.000	0.043	0.038	-0.029	-0.268	0.323	-0.456	0.915	0.010	240
121	5.000	1.000	3.000	1.000	-0.371	-0.196	-0.223	2.618	1.828	-0.133	-0.504	0.759	137
122	5.000	2.000	3.000	1.000	-0.166	-0.032	-0.296	1.356	-0.232	-0.244	-0.318	1.091	199
123	5.000	3.000	3.000	1.000	0.106	0.120	0.148	0.494	0.067	-0.359	0.441	0.193	152
124	6.000	1.000	3.000	1.000	-0.304	-0.470	-0.597	2.496	1.659	-0.395	-0.371	-0.573	144
125	6.000	2.000	3.000	1.000	-0.087	-0.248	-0.510	1.322	-0.326	-0.312	-0.360	-1.230	165
126	6.000	3.000	3.000	1.000	0.262	0.286	0.621	0.717	0.143	-0.299	0.282	0.225	144
127	1.000	1.000	3.000	2.000	-0.197	-0.152	0.196	1.085	-0.046	-0.036	-0.017	0.238	66
128	1.000	2.000	3.000	2.000	-0.036	-0.108	0.142	0.487	0.020	0.056	0.046	0.295	58
129	1.000	3.000	3.000	2.000	-0.000	0.002	-0.002	-0.002	-0.003	0.000	-0.001	0.007	300
130	2.000	1.000	3.000	2.000	-0.113	-0.207	0.152	1.312	0.121	-0.089	-0.185	-0.216	89
131	2.000	2.000	3.000	2.000	0.022	-0.026	0.322	0.591	0.193	-0.083	0.011	0.187	160
132	2.000	3.000	3.000	2.000	0.029	0.020	0.060	0.055	0.094	-0.013	-0.039	-0.468	300
133	3.000	1.000	3.000	2.000	-0.006	-0.370	0.054	1.341	0.337	-0.209	-0.174	-0.404	91
134	3.000	2.000	3.000	2.000	0.146	-0.194	0.198	0.790	0.365	-0.146	0.125	-0.449	141
135	3.000	3.000	3.000	2.000	0.139	-0.104	0.135	0.154	0.165	-0.043	-0.066	-0.429	299
136	4.000	1.000	3.000	2.000	-0.159	-0.209	-0.288	0.914	-0.197	-0.106	-0.262	1.703	63
137	4.000	2.000	3.000	2.000	0.029	-0.053	0.029	0.139	-0.356	-0.336	0.477	0.652	268
138	4.000	3.000	3.000	2.000	0.060	0.055	0.056	-0.232	0.359	-0.445	0.920	0.029	285
139	5.000	1.000	3.000	2.000	-0.290	-0.129	-0.259	2.842	1.273	-0.029	-0.568	0.509	162
140	5.000	2.000	3.000	2.000	-0.056	0.060	0.054	1.693	-0.201	-0.216	-0.306	1.326	210
141	5.000	3.000	3.000	2.000	0.258	0.203	0.088	0.523	0.199	-0.367	0.459	0.171	111
142	6.000	1.000	3.000	2.000	-0.217	-0.418	-0.579	2.456	0.967	-0.347	-0.585	-1.109	153
143	6.000	2.000	3.000	2.000	-0.010	-0.212	-0.367	1.516	-0.260	-0.265	-0.496	-0.818	193
144	6.000	3.000	3.000	2.000	0.432	0.452	0.990	1.033	0.300	-0.261	0.417	0.285	82
145	1.000	1.000	3.000	3.000	-0.166	-0.166	0.029	1.216	-0.092	-0.114	0.100	-0.046	55
146	1.000	2.000	3.000	3.000	-0.041	-0.148	-0.012	0.541	0.085	0.061	0.164	-0.166	36
147	1.000	3.000	3.000	3.000	-0.001	-0.003	-0.009	-0.001	-0.011	0.000	-0.006	0.003	299
148	2.000	1.000	3.000	3.000	-0.100	-0.186	0.117	1.236	0.140	-0.058	-0.162	-0.321	95
149	2.000	2.000	3.000	3.000	0.023	-0.091	0.269	0.717	0.299	-0.059	0.174	-0.298	94
150	2.000	3.000	3.000	3.000	0.027	0.014	0.052	0.053	0.086	-0.013	-0.043	-0.474	296
151	3.000	1.000	3.000	3.000	0.040	-0.376	0.041	1.471	0.346	-0.261	-0.008	-0.241	99
152	3.000	2.000	3.000	3.000	0.157	-0.222	0.258	1.134	0.550	-0.210	0.265	-0.421	60
153	3.000	3.000	3.000	3.000	0.128	-0.097	0.151	0.142	0.218	-0.046	-0.024	-0.418	156
154	4.000	1.000	3.000	3.000	-0.170	-0.069	-0.324	1.165	0.060	-0.167	0.012	1.052	53
155	4.000	2.000	3.000	3.000	0.058	-0.040	0.140	0.220	-0.234	-0.217	0.388	0.667	288
156	4.000	3.000	3.000	3.000	0.073	0.070	0.099	-0.145	0.462	-0.378	0.814	0.031	300
157	5.000	1.000	3.000	3.000	-0.178	-0.131	-0.227	3.032	0.636	-0.096	-0.618	0.210	169
158	5.000	2.000	3.000	3.000	0.047	0.076	0.015	1.794	-0.197	-0.397	-0.162	1.276	159
159	5.000	3.000	3.000	3.000	0.201	0.204	-0.580	-0.247	0.383	-0.343	-0.105	-0.427	126
160	6.000	1.000	3.000	3.000	-0.113	-0.389	-0.537	2.568	0.561	-0.091	-0.684	-1.174	188
161	6.000	2.000	3.000	3.000	0.179	-0.050	-0.142	2.089	-0.135	-0.079	-0.210	-0.464	135
162	6.000	3.000	3.000	3.000	0.441	0.458	0.515	-0.181	0.323	-0.323	-0.166	-0.290	99

Info: Für die 162 Simulationskonfigurationen wurden jeweils nur Replikationen berücksichtigt, welche korrekt als 3-Klassenmodell identifiziert wurden.
Var()=Varianz; KoVar()=Kovarianz; M(Y CX)=Mean des Wachstumsparameters Y der Klasse X

A.3 Mediatoranalysen zum Bias

Tabelle 49: Reg. Erg. zu Bias der Ko-/Varianz Parameter Mediation (robuste SE; HC3)

	<i>Dependent variable:</i>		
	m_bias_i_v_medi	m_bias_s_v_medi	m_bias_is_cv_medi
	Pop.Val.: 0.25	Pop.Val.: 0.04	Pop.Val.: 0.02
	(1)	(2)	(3)
Constant	0.100*** (0.003)	0.020*** (0.001)	0.025*** (0.001)
prob_mean_delta	-0.039*** (0.003)	-0.006*** (0.001)	0.012*** (0.001)
nobs2	-0.013*** (0.001)	-0.003*** (0.0001)	-0.004*** (0.0002)
nobs3	-0.019*** (0.001)	-0.005*** (0.0001)	-0.007*** (0.0002)
mixing2	-0.002** (0.001)	0.0001 (0.0001)	-0.001*** (0.0002)
mixing3	-0.002** (0.001)	0.0004** (0.0001)	-0.002*** (0.0002)
invar2	-0.011** (0.004)	0.001 (0.001)	0.0003 (0.001)
invar3	-0.011** (0.003)	0.004*** (0.001)	-0.001 (0.001)
invar4	-0.003 (0.004)	-0.0001 (0.001)	-0.0004 (0.001)
invar5	0.001 (0.003)	-0.005*** (0.001)	-0.008*** (0.001)
invar6	-0.009** (0.003)	0.004*** (0.001)	-0.009*** (0.001)
sep2	-0.042*** (0.003)	-0.004*** (0.001)	-0.005*** (0.001)
sep3	-0.071*** (0.003)	-0.014*** (0.001)	-0.014*** (0.001)
invar2:sep2	0.007 (0.004)	-0.002** (0.001)	-0.001 (0.001)
invar3:sep2	0.023*** (0.004)	-0.003** (0.001)	0.003** (0.001)
invar4:sep2	-0.007 (0.004)	-0.002** (0.001)	-0.004** (0.001)
invar5:sep2	0.010** (0.004)	0.005*** (0.001)	0.011*** (0.001)
invar6:sep2	0.025*** (0.004)	-0.003*** (0.001)	0.011*** (0.001)
invar2:sep3	0.013*** (0.004)	-0.0005 (0.001)	-0.0001 (0.001)
invar3:sep3	0.031*** (0.003)	-0.002* (0.001)	0.003** (0.001)
invar4:sep3	0.011** (0.004)	0.001 (0.001)	0.0003 (0.001)
invar5:sep3	0.039*** (0.003)	0.010*** (0.001)	0.012*** (0.001)
invar6:sep3	0.085*** (0.003)	0.008*** (0.001)	0.019*** (0.001)
Observations	23,084	23,084	23,084
R ²	0.319	0.348	0.276
Adjusted R ²	0.318	0.347	0.276
Residual Std. Error (df = 23061)	0.033	0.007	0.011
F Statistic (df = 22; 23061)	490.642***	558.669***	400.163***

Info:

*p<0.05; **p<0.01; ***p<0.001

Tabelle 50: Delta-R-Quadrat der Praediktoren zum Bias der Ko-/Varianz Parameter Mediation

	df	dR Var I	dR Var S	dR KoVar IS
prob_mean_delta	1	0.005	0.002	0.004
nobs	2	0.037	0.045	0.049
mixing	2	0.0003	0.0003	0.003
invar	5	0.003	0.030	0.019
sep	2	0.048	0.042	0.022
invar:sep	10	0.068	0.044	0.046

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 49.

Tabelle 51: Reg. Erg. zum Bias der Interceptfaktoren Mediation (robuste SE; HC3)

	<i>Dependent variable:</i>		
	m_bias_i_m_c1_medi	m_bias_i_m_c2_medi	m_bias_i_m_c3_medi
	(1)	(2)	(3)
Constant	0.235*** (0.022)	0.212*** (0.007)	0.222*** (0.017)
prob_mean_delta	1.123*** (0.197)	-0.195*** (0.014)	0.920*** (0.029)
nobs2	-0.038 (0.030)	-0.014*** (0.002)	-0.032*** (0.004)
nobs3	-0.075** (0.027)	-0.021*** (0.002)	-0.053*** (0.004)
mixing2	-0.039 (0.026)	-0.035*** (0.008)	-0.023*** (0.003)
mixing3	-0.094* (0.039)	-0.062*** (0.008)	-0.047*** (0.004)
invar2	0.064* (0.028)	0.002 (0.006)	-0.038 (0.022)
invar3	0.106*** (0.008)	0.013* (0.007)	-0.041 (0.022)
invar4	0.037* (0.016)	0.009 (0.007)	-0.078*** (0.022)
invar5	0.388*** (0.043)	0.191*** (0.009)	-0.282*** (0.018)
invar6	0.391*** (0.039)	0.204*** (0.009)	-0.289*** (0.018)
sep2	-0.029 (0.047)	-0.049*** (0.008)	-0.121*** (0.022)
sep3	-0.143** (0.045)	-0.173*** (0.007)	-0.153*** (0.017)
invar2:sep2		0.024** (0.008)	0.003 (0.027)
invar3:sep2		0.028** (0.008)	0.002 (0.028)
invar4:sep2		-0.017 (0.009)	0.050 (0.027)
invar5:sep2		-0.183*** (0.010)	0.196*** (0.024)
invar6:sep2		-0.151*** (0.011)	0.187*** (0.025)
invar2:sep3		0.012 (0.007)	0.051* (0.022)
invar3:sep3		0.037*** (0.007)	0.039 (0.023)
invar4:sep3		0.131*** (0.008)	0.034 (0.022)
invar5:sep3		-0.102*** (0.009)	0.217*** (0.018)
invar6:sep3		-0.057*** (0.010)	0.196*** (0.019)
mixing2:sep2		0.041*** (0.008)	
mixing3:sep2		0.054*** (0.008)	
mixing2:sep3		0.063*** (0.008)	
mixing3:sep3		0.163*** (0.008)	
Observations	23,084	23,084	23,084
R ²	0.049	0.286	0.264
Adjusted R ²	0.049	0.285	0.263
Residual Std. Error	1.383 (df = 23071)	0.116 (df = 23057)	0.203 (df = 23061)
F Statistic	99.362*** (df = 12; 23071)	355.490*** (df = 26; 23057)	375.830*** (df = 22; 23061)

Info:

*p<0.05; **p<0.01; ***p<0.001

Tabelle 52: Delta-R-Quadrat der Praediktoren zum Bias der Intercept Parameter Mediation

	df	dR Mean I c1	dR Mean I c2	dR Mean I c3
prob_mean_delta	1	0.004	0.010	0.077
nobs	2	0.0004	0.004	0.008
mixing	2	0.001	0.006	0.005
invar	5	0.008	0.088	0.050
sep	2	0.001	0.026	0.006
invar:sep	10		0.076	0.022
mixing:sep	4		0.040	

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 51.

Tabelle 53: Reg. Erg. zum Bias der Slopefaktoren Mediation (robuste SE; HC3)

	<i>Dependent variable:</i>		
	m_bias_s_m_c1_medi	m_bias_s_m_c2_medi	m_bias_s_m_c3_medi
	(1)	(2)	(3)
Constant	0.094*** (0.011)	0.178*** (0.008)	0.136*** (0.009)
prob_mean_delta	0.562*** (0.027)	-0.262*** (0.021)	0.341*** (0.017)
nobs2	-0.006 (0.004)	-0.006* (0.003)	-0.018*** (0.002)
nobs3	-0.008* (0.004)	-0.007** (0.003)	-0.030*** (0.002)
mixing2	-0.025*** (0.003)	0.032*** (0.002)	-0.011*** (0.002)
mixing3	-0.043*** (0.004)	0.070*** (0.004)	-0.016*** (0.002)
invar2	0.010 (0.013)	0.016* (0.008)	0.011 (0.011)
invar3	0.026* (0.013)	0.021** (0.008)	-0.011 (0.010)
invar4	0.030* (0.013)	0.021* (0.009)	0.071*** (0.012)
invar5	-0.074*** (0.011)	0.056*** (0.007)	-0.032*** (0.008)
invar6	-0.083*** (0.011)	0.083*** (0.008)	-0.031*** (0.009)
sep2	0.027* (0.013)	0.067*** (0.009)	-0.036** (0.012)
sep3	-0.060*** (0.011)	-0.159*** (0.008)	-0.102*** (0.009)
invar2:sep2	-0.010 (0.017)	-0.012 (0.012)	-0.023 (0.015)
invar3:sep2	0.031 (0.019)	-0.041*** (0.012)	-0.015 (0.014)
invar4:sep2	0.070*** (0.019)	0.047*** (0.012)	-0.013 (0.016)
invar5:sep2	0.080*** (0.015)	-0.066*** (0.011)	0.081*** (0.013)
invar6:sep2	0.057*** (0.016)	-0.082*** (0.012)	0.065*** (0.013)
invar2:sep3	-0.003 (0.013)	-0.005 (0.008)	0.022* (0.011)
invar3:sep3	0.022 (0.013)	0.007 (0.008)	0.038*** (0.010)
invar4:sep3	0.574*** (0.015)	0.703*** (0.011)	-0.103*** (0.012)
invar5:sep3	0.567*** (0.015)	0.466*** (0.008)	-0.003 (0.009)
invar6:sep3	0.456*** (0.015)	0.365*** (0.009)	0.009 (0.010)
Observations	23,084	23,084	23,084
R ²	0.484	0.675	0.302
Adjusted R ²	0.484	0.674	0.301
Residual Std. Error (df = 23061)	0.223	0.151	0.132
F Statistic (df = 22; 23061)	983.412***	2,172.233***	452.734***

Info:

*p<0.05; **p<0.01; ***p<0.001

Tabelle 54: Delta-R-Quadrat der Praediktoren zum Bias der Slope Parameter Mediation

	df	dR Mean S c1	dR Mean S c2	dR Mean S c3
prob_mean_delta	1	0.017	0.005	0.024
nobs	2	0.0001	0.0001	0.006
mixing	2	0.003	0.009	0.002
invar	5	0.004	0.002	0.007
sep	2	0.001	0.014	0.007
invar:sep	10	0.138	0.233	0.030

Info: Berechnungsgrundlage sind die Modelle aus Tabelle 53.

A.4 Relativer SE-Bias Referenz

Tabelle 55: Referenztabelle: Relativer SE-Bias fuer alle Simulationskonfigurationen (Teil 1)

Confignum	Invar	Sep	Nobs	Mixing	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(I_C3)	M(S_C1)	M(S_C2)	M(S_C3)	N
1	1.000	1.000	1.000	1.000	-0.565	-0.574	-0.451	-0.754	-0.653	-0.701	-0.769	-0.849	-0.806	52
2	1.000	2.000	1.000	1.000	-0.427	-0.598	-0.516	-0.752	-0.574	-0.674	-0.835	-0.816	-0.786	61
3	1.000	3.000	1.000	1.000	-0.014	0.072	0.061	-0.070	0.055	0.043	0.032	0.015	0.108	297
4	2.000	1.000	1.000	1.000	-0.530	-0.554	-0.514	-0.757	-0.682	-0.718	-0.801	-0.836	-0.798	57
5	2.000	2.000	1.000	1.000	-0.414	-0.555	-0.467	-0.707	-0.507	-0.610	-0.797	-0.796	-0.764	75
6	2.000	3.000	1.000	1.000	-0.019	0.078	0.062	-0.070	0.056	0.050	0.037	0.015	0.119	298
7	3.000	1.000	1.000	1.000	-0.563	-0.558	-0.425	-0.731	-0.680	-0.638	-0.764	-0.818	-0.789	58
8	3.000	2.000	1.000	1.000	-0.468	-0.585	-0.550	-0.726	-0.591	-0.729	-0.772	-0.814	-0.859	82
9	3.000	3.000	1.000	1.000	-0.025	-0.365	-0.124	-0.414	-0.269	-0.026	-0.931	-0.667	-0.297	284
10	4.000	1.000	1.000	1.000	-0.432	-0.468	-0.465	-0.744	-0.573	-0.712	-0.791	-0.875	-0.823	57
11	4.000	2.000	1.000	1.000	-0.387	-0.569	-0.449	-0.657	-0.445	-0.645	-0.778	-0.813	-0.839	75
12	4.000	3.000	1.000	1.000	-0.038	0.032	-0.017	-0.104	-0.124	0.020	-0.878	-0.889	0.079	275
13	5.000	1.000	1.000	1.000	-0.421	-0.418	-0.367	-0.495	-0.599	-0.655	-0.712	-0.739	-0.847	90
14	5.000	2.000	1.000	1.000	-0.480	-0.548	-0.489	-0.575	-0.638	-0.693	-0.715	-0.795	-0.844	133
15	5.000	3.000	1.000	1.000	-0.137	-0.131	-0.220	-0.488	-0.356	-0.640	-0.649	-0.815	-0.733	146
16	6.000	1.000	1.000	1.000	-0.433	-0.357	-0.298	-0.560	-0.642	-0.623	-0.597	-0.726	-0.784	92
17	6.000	2.000	1.000	1.000	-0.546	-0.456	-0.478	-0.646	-0.707	-0.687	-0.751	-0.789	-0.821	113
18	6.000	3.000	1.000	1.000	-0.181	-0.098	-0.160	-0.539	-0.567	-0.628	-0.627	-0.858	-0.856	102
19	1.000	1.000	1.000	2.000	-0.498	-0.523	-0.479	-0.745	-0.538	-0.596	-0.816	-0.814	-0.796	58
20	1.000	2.000	1.000	2.000	-0.277	-0.459	-0.412	-0.587	-0.461	-0.671	-0.715	-0.838	-0.805	40
21	1.000	3.000	1.000	2.000	-0.019	0.043	0.089	-0.077	0.004	0.107	0.012	0.058	0.070	285
22	2.000	1.000	1.000	2.000	-0.548	-0.582	-0.514	-0.746	-0.667	-0.682	-0.805	-0.842	-0.803	71
23	2.000	2.000	1.000	2.000	-0.275	-0.442	-0.371	-0.577	-0.461	-0.748	-0.729	-0.804	-0.872	62
24	2.000	3.000	1.000	2.000	-0.025	0.049	0.089	-0.077	0.009	0.116	0.017	0.060	0.082	292
25	3.000	1.000	1.000	2.000	-0.519	-0.579	-0.474	-0.746	-0.649	-0.709	-0.790	-0.825	-0.960	70
26	3.000	2.000	1.000	2.000	-0.249	-0.420	-0.372	-0.612	-0.494	-0.523	-0.805	-0.864	-0.718	74
27	3.000	3.000	1.000	2.000	-0.050	-0.294	-0.094	-0.413	-0.377	0.036	-0.743	-0.572	-0.775	260
28	4.000	1.000	1.000	2.000	-0.515	-0.586	-0.515	-0.758	-0.665	-0.668	-0.800	-0.897	-0.812	49
29	4.000	2.000	1.000	2.000	-0.221	-0.397	-0.420	-0.932	-0.323	-0.664	-0.796	-0.939	-0.837	83
30	4.000	3.000	1.000	2.000	-0.018	0.042	0.025	-0.020	-0.009	0.082	-0.817	-0.874	0.057	288
31	5.000	1.000	1.000	2.000	-0.404	-0.435	-0.410	-0.586	-0.635	-0.705	-0.651	-0.766	-0.866	77
32	5.000	2.000	1.000	2.000	-0.417	-0.512	-0.403	-0.618	-0.574	-0.658	-0.625	-0.769	-0.777	115
33	5.000	3.000	1.000	2.000	-0.067	-0.027	-0.169	-0.557	-0.182	-0.204	-0.948	-0.838	-0.230	153
34	6.000	1.000	1.000	2.000	-0.411	-0.396	-0.326	-0.721	-0.605	-0.657	-0.640	-0.754	-0.765	87
35	6.000	2.000	1.000	2.000	-0.444	-0.403	-0.465	-0.564	-0.699	-0.653	-0.679	-0.796	-0.763	115
36	6.000	3.000	1.000	2.000	-0.146	-0.135	-0.093	-0.723	-0.496	-0.347	-0.755	-0.856	-0.690	77
37	1.000	1.000	1.000	3.000	-0.516	-0.579	-0.581	-0.772	-0.666	-0.710	-0.775	-0.887	-0.812	59
38	1.000	2.000	1.000	3.000	-0.237	-0.420	-0.313	-0.942	-0.575	-0.647	-0.847	-0.969	-0.851	44
39	1.000	3.000	1.000	3.000	-0.057	-0.132	-0.030	-0.722	-0.237	-0.591	-0.727	-0.577	-0.842	182
40	2.000	1.000	1.000	3.000	-0.502	-0.532	-0.527	-0.760	-0.685	-0.643	-0.773	-0.821	-0.825	69
41	2.000	2.000	1.000	3.000	-0.219	-0.357	-0.365	-0.955	-0.575	-0.593	-0.947	-0.839	-0.839	51
42	2.000	3.000	1.000	3.000	-0.063	-0.168	-0.018	-0.751	-0.293	0.030	-0.733	-0.580	-0.846	189
43	3.000	1.000	1.000	3.000	-0.489	-0.522	-0.465	-0.760	-0.649	-0.696	-0.774	-0.856	-0.803	69
44	3.000	2.000	1.000	3.000	-0.189	-0.331	-0.315	-0.726	-0.511	-0.658	-0.797	-0.794	-0.846	51
45	3.000	3.000	1.000	3.000	-0.125	-0.301	-0.119	-0.795	-0.502	-0.496	-0.966	-0.747	-0.940	102
46	4.000	1.000	1.000	3.000	-0.489	-0.523	-0.512	-0.760	-0.676	-0.674	-0.743	-0.805	-0.817	48
47	4.000	2.000	1.000	3.000	-0.215	-0.428	-0.331	-0.738	-0.307	-0.632	-0.910	-0.839	-0.835	120
48	4.000	3.000	1.000	3.000	-0.080	-0.055	-0.045	-0.120	-0.324	-0.604	-0.720	-0.864	-0.838	250
49	5.000	1.000	1.000	3.000	-0.348	-0.357	-0.381	-0.553	-0.622	-0.578	-0.718	-0.769	-0.783	78
50	5.000	2.000	1.000	3.000	-0.314	-0.493	-0.411	-0.691	-0.463	-0.696	-0.723	-0.806	-0.841	123
51	5.000	3.000	1.000	3.000	-0.157	-0.110	-0.192	-0.914	-0.506	-0.534	-0.885	-0.901	-0.520	205
52	6.000	1.000	1.000	3.000	-0.373	-0.317	-0.395	-0.556	-0.534	-0.733	-0.708	-0.752	-0.781	85
53	6.000	2.000	1.000	3.000	-0.344	-0.423	-0.436	-0.709	-0.673	-0.697	-0.766	-0.799	-0.797	110
54	6.000	3.000	1.000	3.000	-0.139	-0.114	-0.142	-0.796	-0.683	-0.424	-0.978	-0.874	-0.786	111
55	1.000	1.000	2.000	1.000	-0.509	-0.550	-0.462	-0.718	-0.590	-0.681	-0.825	-0.974	-0.835	40
56	1.000	2.000	2.000	1.000	-0.398	-0.525	-0.407	-0.924	-0.508	-0.649	-0.914	-0.799	-0.759	74
57	1.000	3.000	2.000	1.000	0.093	-0.064	0.026	-0.080	0.018	-0.011	-0.040	-0.087	0.145	299
58	2.000	1.000	2.000	1.000	-0.501	-0.515	-0.461	-0.716	-0.571	-0.703	-0.791	-0.968	-0.914	44
59	2.000	2.000	2.000	1.000	-0.304	-0.490	-0.422	-0.935	-0.500	-0.638	-0.934	-0.780	-0.687	104
60	2.000	3.000	2.000	1.000	0.090	-0.059	0.025	-0.080	0.017	-0.007	-0.037	-0.092	0.142	298
61	3.000	1.000	2.000	1.000	-0.526	-0.501	-0.427	-0.711	-0.728	-0.661	-0.787	-0.872	-0.808	45
62	3.000	2.000	2.000	1.000	-0.383	-0.474	-0.420	-0.686	-0.501	-0.613	-0.774	-0.819	-0.714	106
63	3.000	3.000	2.000	1.000	0.085	-0.052	0.013	-0.085	0.027	-0.019	-0.031	-0.105	0.123	296
64	4.000	1.000	2.000	1.000	-0.531	-0.535	-0.510	-0.896	-0.610	-0.656	-0.822	-0.793	-0.962	47
65	4.000	2.000	2.000	1.000	-0.222	-0.422	-0.387	-0.551	-0.361	-0.429	-0.718	-0.813	-0.821	170
66	4.000	3.000	2.000	1.000	0.101	-0.072	-0.060	-0.233	-0.254	-0.023	-0.901	-0.909	0.131	263
67	5.000	1.000	2.000	1.000	-0.332	-0.322	-0.310	-0.776	-0.608	-0.692	-0.649	-0.726	-0.796	132
68	5.000	2.000	2.000	1.000	0.056	-0.127	-0.398	-0.477	-0.509	-0.398	-0.217	-0.756	-0.802	194
69	5.000	3.000	2.000	1.000	-0.137	-0.126	-0.197	-0.488	-0.280	-0.079	-0.545	-0.816	-0.356	171
70	6.000	1.000	2.000	1.000	-0.338	-0.317	-0.299	-0.516	-0.626	-0.675	-0.608	-0.726	-0.824	120
71	6.000	2.000	2.000	1.000	-0.400	-0.396	-0.483	-0.606	-0.730	-0.659	-0.668	-0.800	-0.822	143
72	6.000	3.000	2.000	1.000	-0.177	-0.079	-0.141	-0.487	-0.595	-0.744	-0.530	-0.882	-0.591	123
73	1.000	1.000	2.000	2.000	-0.404	-0.413	-0.385	-0.680	-0.615	-0.586	-0.807	-0.803	-0.973	56
74	1.000	2.000	2.000	2.000	-0.164	-0.306	-0.179	-0.852	-0.564	-0.644	-0.762	-0.839	-0.782	58
75	1.000	3.000	2.000	2.000	0.095	-0.025	0.028	-0.096	0.026	0.033	-0.030	-0.005	0.068	299
76	2.000	1.000	2.000	2.000	-0.406	-0.408	-0.360	-0.630	-0.440	-0.603	-0.783	-0.785	-0.800	67
77	2.000	2.000	2.000	2.000	-0.125	-0.265	-0.141	-0.821	-0.547	-0.535	-0.732	-0.794	-0.754	124
78	2.000	3.000	2.000	2.000	0.092	-0.022	0.030	-0.094	0.025	0.042	-0.029	-0.011	0.063	300
79	3.000	1.000	2.000	2.000	-0.386	-0.425	-0.409	-0.826	-0.470	-0.597	-0.772	-0.805	-0.846	64
80	3.000	2.000	2.000	2.000	-0.119	-0.208	-0.146	-0.638	-0.515	-0.634	-0.787	-0.766	-0.634	106
81	3.000	3.000	2.000	2.000	0.072	-0.116	-0.027	-0.129	-0.254	0.012	-0.043	-0.029	-0.044	293

Info: Fuer die 162 Simulationskonfigurationen wurden jeweils nur Replikationen beruecksichtigt, welche korrekt als 3-Klassenmodell identifiziert wurden.
Var()=Var

Tabelle 56: Referenztabelle: Relativer SE-Bias fuer alle Simulationskonfigurationen (Teil 2)

Confignum	Invar	Sep	Nobs	Mixing	Var(I)	Var(S)	KoVar(IS)	M(I_C1)	M(I_C2)	M(I_C3)	M(S_C1)	M(S_C2)	M(S_C3)	N
82	4.000	1.000	2.000	2.000	-0.484	-0.523	-0.500	-0.719	-0.610	-0.696	-0.779	-0.920	-0.816	48
83	4.000	2.000	2.000	2.000	-0.190	-0.378	-0.359	-0.546	-0.314	-0.677	-0.814	-0.836	-0.846	175
84	4.000	3.000	2.000	2.000	0.101	-0.029	0.011	-0.115	-0.151	0.013	-0.790	-0.871	0.066	289
85	5.000	1.000	2.000	2.000	-0.241	-0.348	-0.413	-0.502	-0.585	-0.612	-0.524	-0.734	-0.751	129
86	5.000	2.000	2.000	2.000	-0.398	-0.583	-0.466	-0.571	-0.688	-0.686	-0.626	-0.826	-0.811	183
87	5.000	3.000	2.000	2.000	0.056	-0.062	-0.139	-0.476	-0.250	-0.284	-0.510	-0.877	-0.409	161
88	6.000	1.000	2.000	2.000	-0.333	-0.397	-0.361	-0.539	-0.662	-0.590	-0.639	-0.757	-0.796	110
89	6.000	2.000	2.000	2.000	-0.351	-0.373	-0.455	-0.588	-0.662	-0.565	-0.667	-0.798	-0.739	180
90	6.000	3.000	2.000	2.000	-0.010	-0.085	-0.038	-0.519	-0.434	-0.767	-0.566	-0.884	-0.783	80
91	1.000	1.000	2.000	3.000	-0.541	-0.592	-0.481	-0.731	-0.619	-0.651	-0.791	-0.836	-0.817	54
92	1.000	2.000	2.000	3.000	-0.067	-0.256	-0.101	-0.681	-0.669	-0.813	-0.957	-0.920	-0.892	37
93	1.000	3.000	2.000	3.000	0.082	-0.060	0.059	-0.156	-0.008	0.039	0.049	-0.131	-0.581	281
94	2.000	1.000	2.000	3.000	-0.368	-0.466	-0.393	-0.646	-0.610	-0.563	-0.749	-0.804	-0.768	61
95	2.000	2.000	2.000	3.000	-0.037	-0.260	-0.153	-0.573	-0.619	-0.780	-0.811	-0.982	-0.898	68
96	2.000	3.000	2.000	3.000	0.059	-0.102	0.050	-0.295	-0.157	0.040	-0.025	-0.281	-0.697	277
97	3.000	1.000	2.000	3.000	-0.459	-0.519	-0.409	-0.661	-0.573	-0.653	-0.752	-0.803	-0.799	72
98	3.000	2.000	2.000	3.000	-0.079	-0.264	-0.184	-0.590	-0.607	-0.807	-0.955	-0.849	-0.888	52
99	3.000	3.000	2.000	3.000	0.015	-0.369	-0.081	-0.836	-0.575	0.015	-0.984	-0.734	-0.817	147
100	4.000	1.000	2.000	3.000	-0.547	-0.577	-0.461	-0.873	-0.653	-0.697	-0.828	-0.892	-0.814	41
101	4.000	2.000	2.000	3.000	-0.022	-0.291	-0.238	-0.901	-0.153	-0.662	-0.717	-0.903	-0.831	230
102	4.000	3.000	2.000	3.000	0.073	-0.062	0.017	-0.007	-0.198	0.030	-0.767	-0.898	0.070	297
103	5.000	1.000	2.000	3.000	-0.260	-0.339	-0.384	-0.471	-0.615	-0.682	-0.656	-0.747	-0.906	145
104	5.000	2.000	2.000	3.000	-0.294	-0.510	-0.413	-0.884	-0.487	-0.586	-0.834	-0.860	-0.813	160
105	5.000	3.000	2.000	3.000	-0.032	-0.063	-0.245	-0.851	-0.682	-0.452	-0.969	-0.917	-0.458	190
106	6.000	1.000	2.000	3.000	-0.324	-0.300	-0.337	-0.504	-0.544	-0.660	-0.663	-0.822	-0.791	145
107	6.000	2.000	2.000	3.000	-0.297	-0.344	-0.397	-0.610	-0.512	-0.673	-0.929	-0.774	-0.823	121
108	6.000	3.000	2.000	3.000	-0.010	-0.119	-0.155	-0.928	-0.767	-0.258	-0.981	-0.884	-0.439	110
109	1.000	1.000	3.000	1.000	-0.408	-0.486	-0.546	-0.716	-0.596	-0.611	-0.679	-0.868	-0.763	62
110	1.000	2.000	3.000	1.000	-0.234	-0.360	-0.329	-0.954	-0.547	-0.603	-0.895	-0.783	-0.787	91
111	1.000	3.000	3.000	1.000	0.008	-0.018	0.078	-0.026	-0.022	-0.007	0.019	-0.098	0.106	300
112	2.000	1.000	3.000	1.000	-0.472	-0.530	-0.565	-0.712	-0.682	-0.638	-0.768	-0.912	-0.768	60
113	2.000	2.000	3.000	1.000	-0.218	-0.309	-0.333	-0.931	-0.526	-0.614	-0.756	-0.752	-0.820	141
114	2.000	3.000	3.000	1.000	0.006	-0.016	0.079	-0.023	-0.018	-0.005	0.029	-0.101	0.105	300
115	3.000	1.000	3.000	1.000	-0.449	-0.514	-0.498	-0.732	-0.615	-0.710	-0.738	-0.854	-0.813	64
116	3.000	2.000	3.000	1.000	-0.168	-0.358	-0.257	-0.703	-0.485	-0.401	-0.970	-0.954	-0.731	116
117	3.000	3.000	3.000	1.000	0.007	-0.011	0.074	-0.026	-0.022	-0.004	0.036	-0.096	0.087	300
118	4.000	1.000	3.000	1.000	-0.516	-0.527	-0.480	-0.878	-0.516	-0.748	-0.852	-0.804	-0.864	60
119	4.000	2.000	3.000	1.000	-0.132	-0.286	-0.122	-0.453	-0.138	-0.102	-0.545	-0.819	-0.828	243
120	4.000	3.000	3.000	1.000	0.016	-0.025	-0.025	-0.356	-0.380	-0.011	-0.850	-0.864	0.116	240
121	5.000	1.000	3.000	1.000	-0.198	-0.205	-0.279	-0.542	-0.591	-0.666	-0.621	-0.675	-0.823	137
122	5.000	2.000	3.000	1.000	-0.210	-0.293	-0.456	-0.293	-0.482	-0.456	-0.501	-0.828	-0.844	199
123	5.000	3.000	3.000	1.000	-0.082	-0.027	-0.153	-0.365	-0.153	-0.047	-0.475	-0.623	0.092	152
124	6.000	1.000	3.000	1.000	-0.305	-0.268	-0.289	-0.826	-0.665	-0.613	-0.720	-0.713	-0.820	144
125	6.000	2.000	3.000	1.000	-0.303	-0.383	-0.481	-0.637	-0.811	-0.632	-0.582	-0.823	-0.860	165
126	6.000	3.000	3.000	1.000	-0.105	-0.031	-0.021	-0.388	-0.542	-0.359	-0.442	-0.900	-0.368	144
127	1.000	1.000	3.000	2.000	-0.438	-0.498	-0.502	-0.686	-0.568	-0.662	-0.732	-0.819	-0.775	66
128	1.000	2.000	3.000	2.000	-0.116	-0.265	-0.103	-0.730	-0.689	-0.783	-0.669	-0.831	-0.869	58
129	1.000	3.000	3.000	2.000	0.007	0.009	0.063	-0.040	-0.019	-0.039	0.036	-0.020	0.037	300
130	2.000	1.000	3.000	2.000	-0.397	-0.502	-0.423	-0.664	-0.480	-0.621	-0.721	-0.795	-0.774	89
131	2.000	2.000	3.000	2.000	-0.121	-0.246	-0.104	-0.693	-0.593	-0.755	-0.709	-0.784	-0.857	160
132	2.000	3.000	3.000	2.000	0.006	0.008	0.065	-0.033	-0.017	-0.030	0.045	-0.022	0.032	300
133	3.000	1.000	3.000	2.000	-0.474	-0.524	-0.525	-0.715	-0.640	-0.702	-0.764	-0.823	-0.852	91
134	3.000	2.000	3.000	2.000	-0.113	-0.176	-0.163	-0.677	-0.467	-0.617	-0.750	-0.762	-0.628	141
135	3.000	3.000	3.000	2.000	0.009	-0.038	0.055	-0.063	-0.101	-0.027	0.042	-0.026	0.023	299
136	4.000	1.000	3.000	2.000	-0.500	-0.565	-0.491	-0.762	-0.460	-0.663	-0.813	-0.915	-0.808	63
137	4.000	2.000	3.000	2.000	-0.035	-0.198	-0.183	-0.475	-0.199	-0.668	-0.609	-0.818	-0.839	268
138	4.000	3.000	3.000	2.000	0.029	-0.005	0.006	-0.235	-0.186	-0.057	-0.769	-0.868	0.036	285
139	5.000	1.000	3.000	2.000	-0.284	-0.327	-0.397	-0.479	-0.626	-0.594	-0.461	-0.618	-0.810	162
140	5.000	2.000	3.000	2.000	-0.317	-0.528	-0.461	-0.174	-0.561	-0.638	-0.420	-0.847	-0.864	210
141	5.000	3.000	3.000	2.000	0.003	-0.054	-0.169	-0.571	-0.296	-0.298	-0.579	-0.888	-0.533	111
142	6.000	1.000	3.000	2.000	-0.240	-0.257	-0.236	-0.780	-0.627	-0.559	-0.624	-0.720	-0.909	153
143	6.000	2.000	3.000	2.000	-0.256	-0.335	-0.449	-0.507	-0.699	-0.513	-0.404	-0.812	-0.774	193
144	6.000	3.000	3.000	2.000	0.027	-0.096	0.020	-0.437	-0.327	-0.198	-0.531	-0.878	-0.215	82
145	1.000	1.000	3.000	3.000	-0.384	-0.518	-0.558	-0.602	-0.618	-0.657	-0.655	-0.862	-0.976	55
146	1.000	2.000	3.000	3.000	-0.064	-0.244	-0.035	-0.751	-0.675	-0.477	-0.824	-0.863	-0.974	36
147	1.000	3.000	3.000	3.000	0.032	-0.005	0.041	-0.065	-0.030	-0.021	0.101	0.071	0.070	299
148	2.000	1.000	3.000	3.000	-0.415	-0.497	-0.449	-0.719	-0.503	-0.611	-0.782	-0.852	-0.768	95
149	2.000	2.000	3.000	3.000	-0.052	-0.211	-0.062	-0.671	-0.637	-0.608	-0.919	-0.798	-0.819	94
150	2.000	3.000	3.000	3.000	0.030	-0.027	0.043	-0.060	-0.175	-0.014	0.111	0.081	0.049	296
151	3.000	1.000	3.000	3.000	-0.370	-0.482	-0.481	-0.715	-0.471	-0.703	-0.752	-0.869	-0.852	99
152	3.000	2.000	3.000	3.000	-0.041	-0.236	-0.008	-0.928	-0.505	-0.752	-0.945	-0.835	-0.900	60
153	3.000	3.000	3.000	3.000	-0.005	-0.401	-0.169	-0.898	-0.676	-0.050	-0.991	-0.786	-0.762	156
154	4.000	1.000	3.000	3.000	-0.461	-0.596	-0.431	-0.685	-0.589	-0.579	-0.732	-0.941	-0.761	53
155	4.000	2.000	3.000	3.000	-0.054	-0.276	-0.082	-0.536	-0.120	-0.775	-0.880	-0.839	-0.887	288
156	4.000	3.000	3.000	3.000	0.019	0.000	0.050	-0.020	-0.063	-0.027	-0.789	-0.926	0.062	300
157	5.000	1.000	3.000	3.000	-0.258	-0.309	-0.332	-0.502	-0.596	-0.558	-0.518	-0.738	-0.783	169
158	5.000	2.000	3.000	3.000	-0.279	-0.441	-0.406	-0.636	-0.344	-0.549	-0.901	-0.889	-0.888	159
159	5.000	3.000	3.000	3.000	-0.138	-0.128	-0.300	-0.903	-0.753	-0.378	-0.988	-0.936	-0.461	126
160	6.000	1.000	3.000	3.000	-0.321	-0.297	-0.227	-0.412	-0.650	-0.616	-0.547	-0.775	-0.757	188
161	6.000	2.000	3.000	3.000	-0.163	-0.299	-0.346	-0.864	-0.611	-0.717	-0.807	-0.875	-0.852	135
162	6.000	3.000	3.000	3.000	-0.145	-0.279	-0.242	-0.962	-0.811	-0.362	-0.988	-0.915	-0.614	99

A.5 Online-Appendix

Im Online-Anhang zu dieser Arbeit finden sich die «Results-Files» der Simulation, welche die Schätzergebnisse der einzelnen Replikationen enthalten. Des Weiteren enthält der Anhang das Template zur Erstellung der einzelnen Inputfiles der Simulation sowie Beispieldateien für die Ansteuerung des BWUniClusters und individuelle Inputfiles. Auf Grundlage dieser Rohdaten können sowohl einzelne (oder alle) Teile repliziert und validiert werden, wie auch weiterführende und modifizierte Analysen aufgebaut werden. Des Weiteren ist im Online-Anhang die visuelle Modelldiagnostik zu den Metamodellen zu Bias, Klassenverzerrung und SE-Bias abrufbar. Die Dateien können unter der Adresse: <https://osf.io/aq29w/> auf der OSF-Plattform (Open Science Framework) des Centers for Open Science (COS) gefunden werden. Auf die Inhalte kann nach folgendem Muster Bezug genommen werden:

- Krause, Thomas. 2019. Appendix Multiple Indicator Growth Mixture Models.
<https://osf.io/aq29w/>.